Computer Simulations - from Peptides to Metalloenzymes

Greisen, Per Junior

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Computer Simulations - from Peptides to Metalloenzymes

by

Per Jr. Greisen

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

TECHNICAL UNIVERSITY OF DENMARK

June 2011

Author .....
Department of Physics
June 6, 2011

Certified by .....
Henrik Bohr
Associate Professor
Thesis Supervisor

Accepted by .....
Jane Hvolbæk Nielsen
Head of Department, Department of Physics

Computer Simulations - from Peptides to Metalloenzymes

by

Per Jr. Greisen

Submitted to the Department of Physics
on June 6, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Structural computational biology is a broad field ranging from quantum mechanical simulation to coarse grained models of cells. In many cases, an atomistic analysis of a problem is necessary to understand the biological observation.
Here, different simulation techniques ranging from quantum mechanical simulation over molecular dynamics simulation to Monte Carlo will be applied depending on the problems to be solved. In this thesis, I focus on the simulations of peptides and metalloenzymes exploring the possibilities with computer simulation.

Thesis Supervisor: Henrik Bohr
Title: Associate Professor

Preface

This thesis is submitted as part of the requirements for the Ph.D. degree from the Technical University of Denmark(DTU). The project was supported by a Ph.D. grant from DTU. The work included in the thesis has been carried out from September 2007 to June 2011 at Department of Physics under the supervision of Associate Professor Henrik Bohr. In between 9 great months of parental leave were squeezed in.
Part of the work present was carried out at the Department of Chemistry, DTU under the supervision of Professor Jens Ulstrup.
For the external stay of the my Ph.D. program(1. August 2009 to 1 May 2010), I was in the Bakerlab at the University of Washington under the supervision of Professor David Baker. The stay was extended by 2 months due to some of the results in Chapter 5.

Outline of Thesis

Chapter 1 is a general introduction to peptides, proteins and metalloenzymes. Furthermore, a short introduction to the field of computational design is given. Chapter 2 reviews the theory behind the methods applied and is mainly meant to give an overview. Chapter 3 describes the computational investigations of the peptide dermorphin and compares it with an isomer of it. Chapter 4 shows the results from simulations of a redox protein in different redox states. Chapter 5 shows repurposing of a metalloenzyme and outlines a general strategy for repurposing of metalloenzymes. Chapter 6 shows computational design of binuclear metalloproteins which have been generated but not yet tested.
Chapters 3 to 6 contain results obtained during my Ph.D studies thesis and in the acknowledgement section my contribution along with each persons contribution to the research. Each chapter contains a method section where all information necessary to continue and finish the projects is available.

Not in the Thesis

During my stay at the Department of Physics and Chemistry at DTU, I have been around many interesting people(Dr. Jens August Lundbæk, Dr. Jesper Søndergaard Hansen, Dr. Claus H. Nielsen and Dr. Kasper P. Kepp) and smaller projects have been done in collaboration with them. The subject of these projects range from membrane physics to copper clusters and I have chosen not to write them into the thesis. Even though, results from these projects led to 1 accepted, 1 submitted, and one manuscript in preparation. The *submitted* and the *manuscript in preparation* are included in the thesis and the submitted articles are attached in the Appendix, but will be removed upon publication and the status list will be updated.

Acknowledgements

There are many people without whom my graduate experience would not as exciting and rewarding as it has been.

Henrik Bohr, my supervisor, only one of a kind in every sense. For showing enthusiasm and giving me freedom to pursue some of my own interest. For helping out when help was needed.

Jens Ulstrup for teaching and showing me that seminal papers and facts should be learn by heart.

David Baker for allowing me to visit the lab and treating me like a member. Furthermore, for fantastic inspiration and enormously enthusiasm and dedication to science. Last but not least for taking me up steep mountains while examining me in my project.

Dr. Karl Jalkanen for nice times at QuP and future projects to be finished. Thanks for good discussions and nice comments and for helping.

Kasper P. Kepp for guiding me in the right direction between ordered zinc cluster and distorted copper cluster with the density as the only weapon. Great discussion on DFT and copper.

Claus H. Nielsen and Jesper S. Hansen for interesting discussions on membrane proteins and their function plus great times in Boston and San Francisco.
Jens August Lundbæk - a troubadour in science with whom I have had many great conversation and fun. He is a big inspiration showing great dedication to science. Teaching me the importance of keeping deadlines and to never do important work in the last minute. The only problem is that I will never be interested in membranes.
Ph.D. students at DTU physics: Jacob Laugesen, Peter Larsen and Christian Bentzen - Anders Andersen(well in many ways one of the students) for making me believe that my Elo number should be greater than 300.
Florian Richter for helping me getting started with Rosetta and implementing code.
The Bakerlab for being nice to me - especially Possu Huang and Ingemar Andre for showing me the fantastic nature in USA.
My fellows in crime from the Bakerlab: Sarel Fleishman and Nobuyashu Koga, two extremely bright and surely will do great science in the future. It was fantastic sharing the office with you guys.

Dr. Sagar Khare for good discussion on designs but too much indian time.

Above all I would to thank the following people for supporting me and making me who I am today. I would like to thank Mette Greisen for continuously insisting that education matters without being a tigermom and for being very helpful when needed. My sister Ida, who has helped tremendously during the thesis - always able to pick up my children when needed. You have been amazing. Last but not least my little family - Olga - my beloved girlfriend, half of who I am, for keeping up with me during the ups and downs during the thesis. You have always been supportive and understand keeping the whole family together - the love of my life. To Johanna and Alfred for always making me smile and making life wonderful.

Summary

To understand the relationship between the structure and the activity of peptides or how enzymes catalyze chemical reactions is of utmost importance in life sciences.
One study is about peptides that are smaller entities compared to proteins. They have many different functions, ranging from antimicrobacterial to neurotransmitters. Peptides have been isolated as signal transmitters in the brain binding to the opiate receptor. From South American amphibians, a peptide has been isolated, dermorphin, which binds specifically to a subtype of the opiate receptor. The peptide contains a modification at the second position where it contains a D-isomer of the amino acid alanine. By reverting the D-isomer back to the L-isomer binding of the peptide to the opiate receptor is weak. The two isomers are simulated to investigate their difference in structural flexibility. It is observed that there is a decreased flexibility in the N-terminal part of the peptide with the D-alanine.
Another study is on the enzyme, nitrite reductase, in the nitrogen cycle, which is vital for life as we know it. Nitrite is part of the denitrofixation process and can be either oxidized or reduced by the copper containing enzyme, nitrite reductase. To reduce nitrite, it is necessary to transfer an electron from one part of the protein to the active site of the enzyme where nitrite is reduced to nitrogen monoxide. Here, the electronic pathways in different redox states are investigated and by monitoring thermal fluctuations two pathways have been identified.
The third study is on enzyme catalysis and designing new enzymes. The potential of using enzymes in different processes is enormous, although, not many enzymes are available for synthetic purposes. Lately, computational enzyme design has shown great results, nevertheless, there is room for improvements. Here, mononuclear zinc

enzymes are redesigned to catalyse the hydrolysis of a phosphotriester reaction. Potential candidates are computationally created and later tested in the laboratory where one has shown activity. After the success with a mononuclear metal site - why not try a binuclear metal site. Computational designs are generated to hydrolyse a phosphotriester and promising candidates are created, but have not yet been tested experimentally.

The thesis shows that by using computer simulations to supplement experimental data, it is possible to access states which can be hard or impossible to determine experimentally. Furthermore, reversing the knowledge obtained, one can predict changes and make new enzymes for new reactions. Still, much more understanding is needed in the field of computational biology and many problems still have to be solved to get catalytic rates similar to native enzymes.

Resume

At forstå de strukturelle ændringer som afgør om et peptid binder til en receptor eller hvordan et enzyme katalyserer en reaktion har stor betydningen for vores forståelse af biologiske processer. Denne viden kan omsættes til at designe biologiske systemer, som kan erstatte ikke-vedvarende teknologier.
Proteiner og peptider er de redskaber, som en celle benytter sig af, når den skal udføre arbejde. De er opbygget af aminosyrer, hvor peptider er defineret som kortere polymere samlignet med proteiner. Sammensætningen af aminosyrer giver dem deres specifikke struktur og funktion. Peptider har mange funktioner og kan fungere lige fra antimikrobakterielle til neurotransmitter. Peptider er blevet isoleret i hjernen, hvor de binder til opiatreceptoren. Fra sydamerikanske frøer er isoleret peptidet, dermorphin, som binder specifikt til en bestemt undertype af opiatreceptoren. Peptidet har en isomer substitution på dens anden position og indsættes den mere almindelige L-isomer af aminosyren på denne position, er peptidet ikke længere i stand til at binde til receptoren med samme specificitet. Via molekylær dynamiske beregninger af de to isomere ses en nedsat fleksibilitet af det bindende peptide, som indeholder isomer substitutionen.

Nitrogen fikseringen er grundlaget for liv og bruges i proteiner samt DNA. Nitrite er en del af denne process og kan enten oxideres eller reduceres videre i denne cyklus. Nitrite reduceres er et enzym, som indeholder to kobber ioner. For at reducere nitrite er det nødvendigt at overføre en elektron fra denne ene kobber ion til den anden, som er i det aktive site. Forskellige elektroniske veje er blevet undersøgt og det viser sig, det er muligt at overføre elektronen ad forskellige veje afhængigt af termiske

fluktuationer.

Enzymer katalyserer kemiske reaktioner ved at nedsætte aktiveringsenergien, således at reaktioner kan forløbe. Potentialet for at benytte enzymer til forskellige processer er uanede, men desværre findes ikke enzymer til alle syntetiske reaktioner. At kunne ændre enzymer til specifikke reaktioner har et stort potentiale. Det er blevet vist, at det er muligt at lave nye enzymer via computer simuleringer, men deres ratekonstanter er langt fra naturligt forekommende enzymer. En måde at forøge ratekonstanterne på kunne være at bruge metalioner, som er til stede i mange naturlige enzymer, hvor specielt zink er brugt. I dette studie har det været muligt at omdanne et zink enzym til at hydrolysere en fosfotrister ved hjælp af computer simuleringer.
Da successen med en metal ion i et enzym lykkedes, hvorfor så ikke forsøge med to metalioner. Computerdesign af enzymer med to metal ioner er blevet lavet til at hydrolysere en fosfotriester. De nye sekvenser er lavet, men er endnu ikke blevet testet eksperimentielt.

Afhandlingen viser at computere kan bruges til at supplere eksperimentielle data og give en bedre forståelse for de mekanismer, som bruges i naturen. Ved at opnå et højere kendskab til disse processer, er det muligt at bruge dennne viden til at lave biologiske redskaber såsom nye enzymer.

List of Article and Manuscripts

List of Publications and Manuscripts produced during thesis

- Greisen, P Jr., Nielsen, B. G., Bohr, H. N-terminal Flexibility is Decreased By D-isomer Substitution in Dermorphin, *Submitted to Chemical Biology and Drug Design*

- Hansen, J., Vararattanavech, A., Plasencia, I., Greisen, P. Jr., Bomholt, J., Torres, J., Emneus, J., Nielsen, C. H., Interaction between sodium dodecyl sulfate and membrane reconstituted aquaporins: a comparative study of spinach SoPIP2;1 and E. coli AqpZ, *Accepted for publication in Biochimica et Biophysica Acta - Biomembranes*

- Greisen, P. Jr., Lum K, Ashrafuzzaman M.D., Greathouse D.V., Andersen O.S., Lundbæk J.A., LINEAR RATE-EQUILIBRIUM RELATIONS ARISING FROM CHANNEL-BILAYER ENERGETIC COUPLING Amphiphile regulation of gramicidin channel gating, *Submitted to Proceedings of the National Academy of Sciences*

- Greisen, P. Jr., Jespersen, J. B., Kepp, K. P., The Building Blocks of Metallothioneins II: Zn^{2+}- and Cu^{2+}-Clusters from First-Principles Calculations, *Manuscript In Preparation*

Contents

List of Figures

List of Tables

List of Abbreviations

A Acceptor

ADA Adenine Deaminase

Ala Alanine

Asp Aspartic Acid

BO Born-Oppenheimer

BCP Blue copper proteins

CB Covalent Bond

Cys Cysteine

CG Conjugated Gradient

Cu Copper

D Donor

DE Direct Evolution

DFT Density Functional Theory

DM Dermorphin

ET Electron Transfer

esp Electrostatic Potential

Fe Iron

GB Generalized Born

GGA Generalized Gradient Approximation

Gln Glutamine

Glu Glutamic acids

Gly Glycine

H Hamiltonian

hydroxyl the oxygen of the hydroxyl anion

HB Hydrogen bond

Ile Isoleucine

IR Infrared

KS Kohn-Sham

LDA Local Density Approximation

Leu Leucine

L-DM [L-Ala^2]-Dermorphin

LJ Lennard-Jones

Lys Lysine

Met Methionine

Mg Magnesium

Nir Nitrite Reductase

P1 Pathway 1

P2 Pathway 2

Phe Phenylalanine

phophoryl the oxygen of the phosphoryl anion

Pro Proline

PTE Phosphotriesterase

PT Phosphotriesters

QM Quantum mechanics

rdf Radial Distribution Function

REMD Replica Exchange Molecular Dynamics

rot Rotamer

RU Rosetta Units

S Through Space

SD Steepest descent

Ser Serine

SOD Superoxide Dismutase

SS Secondary Structure

T1 Type 1

T2 Type 2

T_D Tetrahedral

T_{BP} Trigonal Bipyramidal

Thr Threonine

Trp Tryptophan

Tyr Tyrosine

Val Valine

WT Wild type

XC Exchange-correlation

Chapter 1

Introduction

1.1 Peptides

Peptides are small entities built by amino acids and can consist of as many residues as 50. They have a wide range of different functions ranging from neurotransmitters[75] to antimicrobial effectors[45]. The design of proteins and peptides is one of the most challenging research topics in the life sciences today. Peptides are also the focus here since they are more realistic targets for design than proteins. Previously, the development of compounds has mainly been focusing on chemical synthesis, but peptide-based therapeutics is a growing field of great interest[142]. The conformational space for peptides is large e.g. for heptapeptides, which are considered small entities, the combinatorial just in sequence space is 20^7 which is around one billion possibilities. The complexity of the problem requires or necessitates approximations and many first principle physics based methods cannot be used in the first round to narrow down the phase space. The methods have been extended by including machine learning approaches such as neural networks. Here, not only structural data is included, but also physico-chemical parameters such as hydrophobicity, size, polarity, etc[119]. The computational methods have demonstrated some success in the design of oseltamivir[135] or in the prediction of anti-microbial peptides[88].

1.1.1 Stabilization of Peptides

A practical problem is that the stability of the peptides is very low and they can easily be hydrolysed by proteases[77]. To enhance their stability, posttranslational modifications are added to the peptides which increase their lifetimes.
Many peptides are protected by capping their N- or C-terminal with either an amidation of the C-terminal or an amino-acetylation at the N-terminal. This prevents the enzymatic degradation of the peptides and increases their lifetimes.
Another modification is a substitution of the natural L-isomer of the amino acids with its D-enantiomeric form. This modification has implications for the dynamics of the peptide and can change the properties of the polymer[1]. The modifications are observed *in vivo* and prolong duration of specific peptides.

1.2 Proteins

Proteins are bigger entities compared to peptides and function as machines in living cell. To function, most proteins have to fold into well defined three dimensional structures which are determined by their sequence and the environment in which they are formed. The fold of the protein is predetermined which has been established by Levinthal[118]. A random exploration of the configurational space would require more time than the age of the universe.
The three dimensional structure is composed of regular secondary structural elements which are periodically repeated such as an α-helix or a β-sheet. In structural biology, the aim is to correlate the three dimensional structure of the protein with its function to gain an understanding of its purpose.

1.2.1 Metalloproteins

The significance of metal ions in biology is enormous where their uses range from signaling[19] to enzymes[193].
Metals are widely used in proteins, and metalloproteins account for nearly half of all

proteins. The metal ion can be incorporated into the protein either during folding[124] or inserted into the protein after folding[153].
The metals used consist of a large and diverse set ranging from Group I alkali metals e.g., Na^+, to transition metal such as copper(Cu). The interaction between the metal ion and the protein increases the structural complexity, where different coordination groups have to be combined with the protein scaffold.
The metal ions are coordinated in the protein by donating ligand atoms where the most common are sulfur(S), nitrogen(N), or oxygen(O). The ligands have different softness/hardness characteristics, where N is borderline, O is hard, and S is soft. These combinations determine the type of metal binding[145] and the stability of the complex which are empirically defined by the Irving-Williams series[79].

1.2.2 Metalloenzymes

Enzymes accelerate the rates of a chemical reactions by many orders of magnitude lowering the activation energy by a strong binding of the transition state[144, 83]. The major contribution to the catalytic efficiency is due to a preorganized electrostatic active site[199, 200]. Metal ions are often encountered in enzymes due to their catalytic properties.
The diversity of metal ions is large and range from Group II elements like magnesium(Mg) to transition metals like iron(Fe). The use of metal ions can roughly be divided into two:

- Redox processes
- Catalysis

where a rough division estimates that 30 % are redox enzymes while 70 % function in regular catalysis[8].
In redox proteins, the metals are either directly involved in the process of the reaction mechanism like in superoxide dismutase(SOD) or in the transfer of an electron like the Type 1 site in Nitrite Reductase(Nir). In non-redox active complexes, the metal

ion functions by polarizing bonds, providing electrostatics stabilization of negative charges, or activating compounds for nucleophilic attack.

1.2.3 The Zinc Ion in Metalloenzymes

Zinc ions, Zn, are often used by eukaryotic enzymes as they were introduced relatively late in evolution[44]. Zn is considered a borderline ion on the Pearson softness/hardness scale. The electronic configuration of Zn is d^{10} which is a closed shell making it redox inactive and stable. The ion resemble the metals found in Alkaline Group II with a radius similar to Mg. It is considered a strong Lewis acid and has very flexible geometry which can be changed during reactions. With four ligands, the coordination of Zn in enzymes is tetrahedral(T_d) as seen in carbonic anhydrase. In case of 5 ligands, it is trigonal bipyramidal(D_{3h}) as in adenine deaminase or square pyramidal(C_{4v}) as in carboxypeptidase.

1.2.4 The Copper Ion in Metalloenzymes

Copper, Cu, is a transition metal and can have two different oxidation states, +1 or +2.

The Cu^{2+} is a borderline between the soft and hard Lewis acid classification, while Cu^{+} is soft. The Cu^{+} has coordinations ranging from 2 to 6, while Cu^{2+} has coordination ranging from 4 to 6[8]. In the Cu^{2+} state, the ion has a d^9 electron configuration which gives it an unpaired electron giving a spin. Due to the degeneracy of the energy levels in Cu^{2+}, it is possible to have symmetry breaking known as the Jahn-Teller effect giving distorted copper sites. Cu^{2+} is mainly T_d coordinated as in Pseudoazurin, or square planar(D_{4h}) like in Cu-Zn SOD. With 5 coordinating entities the D_{3h} is observed as in azurin II[160].

In the reduced state, Cu^{+} has a d^{10} electron configuration with no unpaired electrons. Both redox states are found in nature and copper is involved in redox processes either in connection with electron transfer or as a radical scavenger. For Cu^{+}, the main coordination numbers are 4,5, or 6, but 2 and 3 have also been observed[71].

1.3 Computational Enzyme Design

Enzymes are used in many applications due to their large acceleration of reaction rates. There are still many synthetic reactions that do not have enzymes that are able to catalyze their reactions. The naturally occurring enzymes are perfectly designed for cellular processes refined through evolution, while artificially designed enzymes are optimal for the conditions of industrial processes. Furthermore, the k_{cat} of natural enzymes is not necessarily optimized for a maximal turn-over rate[14] and rational design can be used to improve these values in many cases.
Directed evolution(DE) is a method based on randomness where evolutionary pressure is applied to enzymes *in vitro*. It has generated enzymes with new functions[20].
Computational design of enzymes is a way to rationalize the development of enzymes with new functions. Many different algorithms have been developed to solve the problem. Recently, the method has shown some success where designed enzymes have been able catalyse the three classical chemical reaction types:

- Kemp elimination[159]
- Retro-aldol[89]
- Diels-Alder[176]

The initial chemical rates have been low but in combination with DE, the catalytic rates of the designed enzymes have started to approach those of catalytic antibodies. The catalytic rates are still far from the reaction rates observed for naturally occuring enzymes[13].
Other approaches have been applied in the design of enzymes with new functions. Here, metal ions have been incorporated into the protein to take advantage of their reactivity[65]. For redox reactions, the use of metal ions is necessary to donate or accept electrons during the chemical reaction.

1.3.1 Design of Metalloenzymes

The de novo design of metalloproteins is very challenging, but some success has been achieved. In fact, α-helical proteins have been designed which bind haem-groups and are able to carry oxygen[99, 156]. Protein-protein complexes have also been designed through the coordination of metal ions at the interface[165].

In the design of metalloproteins, computer algorithms have been used to rationalize the process such as METAL SEARCH[36] and DEZYMER[66] where a binding site for iron has been made[150].

1.4 Questions Asked in the Study

An overall aim in the following study is to use computers to gain a deeper insight and understanding into the important processes at an atomistic level. Computers are able to decompose the energy function into the different contributions and determine their relative contributions.

It is difficult to determine the binding conformation of peptides to either a receptor or a membrane. In chapter 3, the conformational space of a peptide along with its modification will be investigated in a case where it is known that the L-isomer of alanine binds much weaker compared with its D-isomer.

Electron transfer is used in many redox processes and in photosynthesis. To determine the electronic pathways and how much different parts of a protein contribute to the electron transfer process is hard to determine experimentally. In chapter 4, electron transfer in Nitrite Reductase is studied computationally to investigate the pathways involved in the transfer. Furthermore, the different redox states of the enzymes are investigated.

The ability to manipulate and change enzyme function is of vital importance. In chapter 5, the repurposing of metalloenzymes is performed using computational techniques and in chapter 6, computational design of binuclear metalloproteins is performed.

Chapter 2

Theory

2.1 Density Functional Theory

For a more comprehensive description of density functional theory see Martin[129] or Koch and Holthausen[98].

2.1.1 Kohn-Sham Orbitals and Equations

The basic idea of density functional theory(DFT) is that the ground state energy of an electronic system can be written in terms of the electron density,ρ. The electronic energy is a functional of the electron density and is denoted E[ρ] in the sense that for a given function, ρ(R), there is a single corresponding energy.

Hohenberg and Kohn[69] proved that the ground state energy is uniquely determined by the electron density. From the theorem, it is not possible to deduce how the energy functional depends on the electron density of the system. The theorem and its proof only states there is one such functional.

Kohn and Sham[100] showed that the exact ground state energy of an n-electron

system can be written as

$$E[\rho] = -\frac{\hbar}{2m_e}\sum_{i=1}^{n}\int \psi_i^*(r_1)\nabla_1^2\psi_i(r_1)dr_1 - \sum_{i=1}^{N}\int \frac{Z_I}{r_{I1}}\rho(r_1)dr_1 + \frac{1}{2}\int\int \frac{\rho(r_1)\rho(r_2}{r_{12}}dr_1dr_2 + E_{EX}[\rho] \tag{2.1}$$

where the one-electron spatial orbital,ψ_i are the the Kohn-Sham(KS) orbitals. The first term is the kinetic energy of the electron and the second term is the electron-nuclear attraction which is summed over all nuclei, N_I, and their charge, Z_I. The third term is the Coulomb interaction between the total electron charge distribution at r_1 and r_2. The last term is the exchange-correlation(XC) energy of the system where all non-classical interactions are computed and is a functional of the density. The exact electronic ground state density is given by

$$\rho(r) = \sum_{i=1}^{n}|\psi_i(r)|^2 \tag{2.2}$$

where the sum is over all occupied KS orbitals and hence ρ is known once all the orbitals have been determined.

The KS orbitals are found by solving the KS equations which are derived by applying the variational principle to the electronic energy E[ρ]. The KS equations for one-electron orbitals $\psi_i(r_1)$ have the following form

$$-\frac{\hbar^2}{2m_e}\nabla_1^2 - j_0\sum_{i=1}^{N}\frac{Z_I}{r_{I1}} + j_0\int\frac{\rho(r_2)}{r_{12}}dr_2 + V_{XC}(r_1)\psi_i(r_1) = \epsilon_i\psi_i(r_1) \tag{2.3}$$

where ϵ_i is the KS orbital energy, V_{XC} is the functional derivative of the XC energy with respect to the density

$$V_{XC}[\rho] = \frac{\delta E_{xc}[\rho]}{\delta\rho} \tag{2.4}$$

and if $E_{xc}$is known V_{XC} can be obtained. The equations are solved in an iterative and self-consistent manner.

2.1.2 Exchange-Correlation Functionals

The main problem in DFT is that the functional for V_{XC} and E_{XC} are not known. As no exact functional exists to describe the electron-electron interaction, different functionals have been developed. They all contain approximations and it is necessary to known the limitations of the functionals.

In general, functionals can be divided into three main groups which use different approximations for the XC: local density approximation(LDA), generalized gradient approximation(GGA), and meta hybrid functionals which also depend on the kinetic energy as well as the incorporation of exact exchange, hybrid functionals.

Here, three functionals have been used for geometry optimization and hessian calculations: BP86[15, 146], B3-LYP[185, 16, 115, 114], and TPSSH[148, 183]. The BP86 functional is GGA and approximates both the exchange as well as the correlation. It does not include any exact exchange(0 % exact exchange). B3LYP Functional is a hybrid functional and uses 20 % of exact exchange. TPSSH Functional is a meta hybrid functional and uses 10 % of exact exchange[147, 148].

2.2 Electron Transfer Theory

The field of charge transfer in biological systems is large and much research has been performed. For a more in depth description of electron transfer in biological system is referred to the seminal book on the subject by Kuznetsov and Ulstrup[106] or one of the following reviews by Marcus[126], Harry Gray[61], or Marcus and Sutin[127].

2.2.1 Long-Range Electron Transfer

In proteins, electrons often are transferred over relatively long distances, > 10 Å, which is considered long-range electron transfer(ET)[60]. In long-range ET, the coupling between the two redox centers is generally weak for intra- and intermolecular transfer.

The process is fast and the Born-Oppenheimer approximation is applied to much of

the theory describing ET where the motions of the electrons and the nuclei are separated. This is in some cases extended to the Franck-Condon principle which further states that during the fast motion of the electrons, the nuclei do not have time to change their position or momenta.

In the weak coupling regime, the donor(D) and acceptor(A) have little or no electronic overlap, H_{AB}, which is much smaller than thermal energy at room temperature

$$H_{AB} \ll k_B T$$

where k_B is the Boltzmann constant and H_{AB} is the electronic coupling strength. This makes long-range ET a non-adiabatic process, where many attempts are necessary for the transfer of the electron to occur.

The exchange mechanism can in general proceed either through a superexchange or a hole mechanism.

In biological systems, superexchange has been observed in redox proteins[29] where the conducting orbitals are higher in energy compared to the donor or acceptor orbitals. In the hole mechanism, the energy levels are reversed as is seen in DNA[191]. It has been shown by Marcus[125] that ET not only is dependent on the free energy, but also on the outer sphere response to the process as well as the inner shell vibrations which are described by the reorganization energy of the process. This leads to a semiclassical expression for the rate constant which can be written as

$$k_{ET} = \frac{2\pi}{\hbar} * H_{AB}^2 * \frac{1}{\sqrt{4\pi\lambda RT}} \exp(-(\Delta G^0 + \lambda)^2/(4 * \lambda RT))$$

where Δ G^0 is the driving force, λ is the reorganization energy, and H_{AB}^2 is the square of the electronic coupling as mentioned before. This expression reduces the complex problem of all the nuclear configurations into an expression in terms of H_{AB}^2 and λ.

2.3 Molecular Dynamics Simulation

All the dynamical properties could in principle be deduced by solving the time dependent Schrödinger equation which contains all the information of the system. The time-dependent Schrödinger equation can be solved for around 100 atoms depending on the level of accuracy required.

Applying the BO approximation the motions of the electrons and nuclei are separated which reduces the complexity. The motion of the nuclei can be described using Newton's equations of motion and it is possible to simulate millions of atoms and reach millisecond time scale[172, 39].

2.3.1 Hamilton's Equations of Motion

Starting in the microcanonical ensemble with a fixed number of particles, N, a confined volume, V, and the total conserved energy, E, the Hamiltonian(H) of the system is written as

$$H = T + V \tag{2.5}$$

where T is the kinetic energy and V is the potential energy of the system. The time evolution of the Hamiltonian is known as Hamilton equations of motion

$$\dot{p} = -\frac{\partial H}{\partial q} \qquad \dot{q} = \frac{\partial H}{\partial p} \tag{2.6}$$

where $\dot{p}$ is the time derivative of the momenta and $\dot{q}$ is the time derivative of the generalized coordinates. The Hamilton's equations of motion can be transformed into Cartesian coordinates and related to the force using Newton's second law written as

$$F_i = m_i * \frac{dv_i}{dt} = m_i * \frac{d^2 r_i}{dt^2} \tag{2.7}$$

where F is the force acting, v_i is the velocity, m_i is the mass of the particle, and r_i is the position vector. The same can be written for the potential energy

$$F_i = \nabla V_i \tag{2.8}$$

where the force is expressed as the gradient of the potential energy. The equations are combined and velocities are assigned using a Maxwell-Boltzmann distribution. The equations were derived in the microcanonical ensemble and can be transformed into other ensembles by Legendre transformations.

2.3.2 Verlet Algorithm for Integration of Newton's Equations of Motion

Two things have to be fulfilled by the numerical algorithms used to integrate the equations: conservation of energy as well as time reversal. A Taylor expansion for a particle in Cartesian coordinates is performed by its position vector r

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{r}(t) + O(\Delta t^4) \tag{2.9}$$

and the same expansion is performed for the particle's old position

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{r}(t) + O(\Delta t^4) \tag{2.10}$$

the two equations are added together

$$r(t + \Delta t) + r(t - \Delta t) = 2r(t)\frac{f(t)}{m}\Delta t^2 + O(\Delta t^4) \tag{2.11}$$

which gives the Verlet algorithm which preserves the phase space volume and is time reversible.

2.3.3 The Potential Energy Function of Force Field

To simulate biological systems, it is necessary to have a potential energy function that can describe various biomolecules such as proteins, DNA, or RNA. Here, the Amber force field is used where the energy function can be decomposed into two contribution, the bonded and nonbonded interactions

$$\begin{aligned} V_{pot} &= \overbrace{V_{bond} + V_{angle} + V_{dihedral}}^{\text{bonded}} + \overbrace{V_{vdw} + V_{electrostatic}}^{\text{nonbonded}} \\ &= \sum_{bonds} K_r * (r - r_{eq})^2 + \sum_{angles} K_\theta * (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\eta - \gamma)) \\ &\quad + \sum_i \sum_{<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right) \end{aligned}$$

where the bonded terms are composed of a bond which is described by the strength of the bond as a harmonic spring with the force constant K_r, the angle term is harmonic with a force constant K_θ. The dihedral term is a potential function which is dependent on the dihedral angle and phase. This term can have multiple minima and maxima depending on the periodicity of the angle.

The nonbonded terms include van der Waals forces which describes the repulsion and dispersion between atoms. In the Amber force field, they are described using the Lennard-Jones(LJ) potential with r^{-12} as the repulsive part and r^{-6} as the dispersive part. The electrostatics is described by the Coulomb equation where each atom is given a so-called fixed partial charge.

2.4 Rosetta

The Rosetta package is huge and its functionality ranges from NMR[154] to protein design[103]. Here, the most important parts of the full atom energy function is described which are summarized by Rohl el al.[157].

The motivation for a short description of the energy function is that it is based on a

knowledge based potential in contrast to molecular dynamics.

2.4.1 Energy Function in Rosetta

The knowledge-based potential of the energy function consists of the following terminology: torsion potential, van der Waals interaction, hydrogen bonding potential, solvation, residue pair interaction, rotamer energy, reference state, and electrostatics. The **rama**-term is used to describe the torsion potential of an amino acid given its position in a secondary structure element

$$\sum_i -\ln[P(\psi_i, \phi_i | aa_i, ss_i)] \tag{2.12}$$

where ψ_i,ϕ_i are the well-known dihedral angles, aa_i is the type of amino acid, and ss_i is the secondary structure element in which the residue is positioned. The van der Waals is described more or less like in a standard force field

$$\sum_i \sum_{j<i} \begin{array}{ll} \left(\left(\frac{r_{ij}}{d_{ij}}\right)^{12} - 2\left(\frac{r_{ij}}{d_{ij}}\right)^{6} \right) * e_{ij} & \text{if } \frac{d_{ij}}{r_{ij}} > 0.6 \\ (-8759.2\frac{d_{ij}}{r_{ij}} + 5672.0) * e_{ij} & \text{else} \end{array} \tag{2.13}$$

the first part is LJ where e_{ij} is the geometric mean of the atom well depth which is taken from the CHARMM19 force field[22], r is the well depth of the potential.
The hydrogen bond potential is probabilistic

$$\sum_i \sum_j (-\ln[P(d_{ij}|h_j ss_{ij})] - \ln[P(\cos\theta_{ij}|d_{ij}h_j ss_{ij})] - \ln[P(\psi\theta_{ij}|d_{ij}h_j ss_{ij})] \tag{2.14}$$

where d is the acceptor-proton distance, h is a hybridization of the atom(sp^2,sp^3). The potential is also dependent on the secondary structure element, and lastly it is dependent on the angles between the donors and acceptors where θ is the proton-acceptor-acceptor base bond angle(e.g., H–O=C), and the last angle, ψ, is the donor-proton-acceptor bond angle(fx. O-H –N).
The solvation term in Rosetta is based on the Lazardis and Karplus theory[112] and

is given as

$$\sum_i \left[\Delta G_i^{ref} - \sum_i \left(\frac{2\Delta G_i^{free}}{4 * \pi^{3/2} \lambda_i r_{ij}^2} \exp[-d_{ij}^2] V_j + \frac{2\Delta G_j^{free}}{4 * \pi^{3/2} \lambda_j r_{ij}^2} \exp[-d_{ij}^2] V_i \right) \right] \quad (2.15)$$

where λ is the correlation length, V is the atomic volume of the atoms. ΔG^{ref} and ΔG^{free} are the reference and solvated free energy of the atom, respectively, and the solvation energy of atom$_i$ can be written as

$$\Delta G_i^{solvated} = \Delta G_i^{ref} - \sum_{i \neq j} f_i(r_{ij}) V_j \quad (2.16)$$

where the reference solvation energy, $\Delta\ G_i^{ref}$, is the solvation free energy in which the molecule is solvent exposed and the second term is the reduction in solvation in the presence of surrounding groups where the relationship is

$$f_i(r_{ij}) = \frac{2\Delta G_i^{free}}{4 * \pi^{3/2} \lambda_i r_{ij}^2} \exp[-d_{ij}^2] \quad (2.17)$$

The interaction term for polarizable atoms also has knowledge based potential

$$\sum_i \sum_{j>i} - \ln \left(\frac{P(aa_i, aa_j | d_{ij})}{P(aa_i | d_{ij}) P(aa_i | d_{ij})} \right) \quad (2.18)$$

which is a distance dependent way to determine strength of the polar interaction between polar residues besides the hydrogen bond potential.
The rotamers are based on their probability in the Dunbrack library

$$\sum_{j>i} - \ln \left(\frac{P(rot | \psi_i \phi_i) * P(aa_i | \psi_i \phi_i)}{P(aa_i)} \right) \quad (2.19)$$

where rot is a backbone dependent rotamer from Dunbrack.
In enzyme design, it is possible to include the Coulomb term

$$\sum_{i<j} \frac{q_i * q_j}{\epsilon * r_{ij}} \quad (2.20)$$

where ϵ is the dielectric constant and q is the partial charge.

Chapter 3

Neuropeptides

3.1 Introduction

The rational design of drugs is a promising field in research where physicochemical properties or structural information are used to design new lead compounds.
Today, development of new drugs is performed by randomly screening many chemical compounds and to observe if any have the desired function. This is a tedious and very time consuming task and can end without any positive results[40].
Despite the intuitive judgement that the rational drug design would be very productive, the success stories are very limited and only few drugs have been developed through this strategy, e.g. gleevec[41].
The low success rate indicates that it is necessary to have a better understanding of the ligands and the interacting complex. Hence, there is a big need to systematize and improve the process of developing new lead compounds. In the following, the focus is on peptides known to bind to the opiate receptor and their structural relationship.

3.2 Amphibian Peptides bind specifically to Opiate Receptor.

The system of opiate receptors is part of the pain sensing system. It is reasonably characterized and a very studied system. Dermorphin(DM) has been isolated from the skin of the frog, *Phyllomedusa Sauvagei*, and shows high specificity and affinity towards to μ-receptor[132, 133]

To determine specificity and affinity of peptides to the opiate receptors, their binding is tested using the mouse vas deferens(MVD) assay to test for δ-receptor[195] and the guinea-pig ileum(GPI) for μ-receptors[143] affinity and specificity(see Table 3.1 for binding values).

Both peptides contain a chiral inversion at the second position of the amino acid, alanine(Ala), rather than the normal L-alanine. The gene for the amphibian peptides is encoded with the usual codon for the L-isomeric residue Ala. The inversion of the L-DM to the D-isomer is a posttranslation modification of the peptide. It has been shown that the inversion is necessary for both peptides to bind strongly to their receptor[113].

3.2.1 Dermorphin

Besides the chiral substitution at the second position, the peptide contains two aromatic residues in the N-terminal part of the chain which are believed to deliver the "message" of the peptide to the receptor. The C-terminal part of the receptor contain a Tyr, Pro, and Ser at the end of the chain. This part is described as the address of the peptide. Connecting the two parts is a Gly residue which acts as a hinge between the two "domains".

The peptide has been investigated by mutagenesis to see which parts are vital for binding. It has been shown that the Tyr^1 and Phe^3 are necessary for activity while Tyr^5 is less essential[189]. The three last residues in the sequence can be substituted as they only partly influence the binding constant.

Table 3.1: Affinity and Specificity for Ligand Binding to the μ–receptor(Values are IC50 (nM))

	Sequence	MVD	GPI	Ratio(MVD/GPI)
Morphine[21]		150	1215	0.12
Leu^{5}enkephalin[169]	Tyr-Gly-Gly-Phe-Leu	246	11.4	21.6
Met^{5}Enkephalin[102]	Tyr-Gly-Gly-Phe-Met	14.7	157	0.09
Dermorphin[46]	Tyr-D-Ala-Phe-Gly-Tyr-Pro-Ser-NH_2	16.5	1.27	12.99
Dermorphin-OH[46]	Tyr-D-Ala-Phe-Gly-Tyr-Pro-Ser-OH	28.1	4.5	6.26
Dermorphin(1-4)[46]	Tyr-D-Ala-Phe-Gly-NH_2	263.0	27.4	9.60
L-Ala^2-Dermorphin[131]	Tyr-Ala-Phe-Gly-Tyr-Pro-Ser-NH_2	> 15000	> 4000	

3.2.2 Aim of Research

The aim of the following research is to determine if it is possible to detect and quantify a structural difference between the two isomers, [L-Ala^2]-DM(L-DM) and DM, using computational methods.
From generated ensembles of structures, the vibrational spectra are computed using quantum mechanics to supplement Raman and infrared experiments.

3.3 Methods

3.3.1 Molecular Dynamics Simulation of Peptides

The two peptides were generated using *xleap* of AmberTool10 package[26]. The D-alanine was generated from the more abundant L-alanine by changing its internal coordinates. The peptides were minimized for 1000 steps using 500 steps of steepest descent(SD) and 500 of conjugated gradient(CG) in vacuum. The peptides were solvated using the TIP3P[90] model for water with 4738 molecules for the L-DM and 4740 for the DM. After solvation of the system, they were minimized for 2500 steps with 1000 of SD and 1500 of CG. The systems were heated up to 300 K applying a weak restraint on the peptides of 10 kcal/mol for 50 ps. The volumes were fixed to equilibrate the pressure. After equilibration, the pressure was set to 1 bar using isotropic scaling with a relaxation time of 1 ps and the temperature was fixed to 300 K using the Langevin thermostat with a collision frequency of 1 ps^{-1}. Hydrogens were restrained using SHAKE[161] and all non-bonded interactions were cut-off at 10 Å. Full electrostatics was applied using the Particle Mesh Ewald[35] approach with periodic boundary conditions. The MD simulations were performed in the NPT ensemble and the time step was set to 2 fs with the ff99SB force field[76]. The simulations were run for 31 ns where the 1 ns was considered equilibration of the system. From the first runs, 9 new initial configurations were chosen to run giving a total of 300 ns sampling.

3.3.2 Replica Exchange Molecular Dynamics

The sampling of the peptides was enhanced using replica exchange molecular dynamics(REMD)[187, 123]. The peptides were generated with *xleap* of AmberTool10 package[26] with a PBradii of 2. To prevent formation of isomers at high temperature, a chirality restraint was applied on the backbone and gave a total of 272 degrees of freedom for the peptides. The peptides were minimized with the ff99SB force field[76] for 10.000 steps using 5.000 of SD and the rest with CG.

Next, the temperature range was chosen as 2 times the minimum temperature(T_{min}), which was set to 300 K. The number of replicas was chosen in accordance with Fukunishi et al.[54]

$$M \approx \sqrt{dof} * \ln\left(\frac{T_{max}}{T_{min}}\right) \tag{3.1}$$

which gave 12 replicas. The temperature range was chosen as

$$T_i \approx T_{min} * \exp\left(\frac{i}{\sqrt{dof}}\right) \tag{3.2}$$

which gave the following temperatures: 300.0, 318.8, 338.7, 359.8, 382.3, 406.2, 431.6, 458.6, 487.3, 517.7, 550.1, and 584.5 K.

All starting trajectories were equilibrated for 1 ns using the Langevin thermostat with a collision frequency of 1 ps^{-1}. Hydrogens were restrained using SHAKE[161] and the equation of motions were integrated with a step size of 1 fs. The implicit solvation model developed by Onufriev was used[141] and an exchange frequency of 0.5 ps was applied which has been shown to give good convergence[177]. The simulations were run for 150 ns with coordinates being recorded every 4 ps. The exchange frequencies were ≈ 31 % which indicated nice overlap between the potential energies.

Table 3.2: Propensity of Turn Structure

Residue	L-DM[1]	DM[1]	L-DM[2]	DM[2]
Phe[3]	0.17	0.23	0.11	0.16
Gly[4]	0.17	0.22	0.14	0.20
Tyr[5]	0.05	0.15	0.06	0.17

[1] Explicit Molecular Dynamics
[2] Replica Exchange Molecular Dynamics

3.4 Results

3.4.1 Configurational Dynamics of Peptides

To investigate the configurational distributions for the peptides, molecular dynamics(MD) simulations are performed. Three parameters have been used to measure variability between the distribution for the isomers [58]

- Secondary structure element
- End-to-end distance between Cα atoms of Tyr^1 and Ser^7
- Radius of gyration

Secondary Structure Analysis of Isomers

Table 3.2 shows the secondary structure(SS) propensities for the three residues phenylalanine 3, Phe^3, glycine 4, Gly^4, tyrosine 5, Tyr^5. The secondary structural elements are obtained using the algorithm developed by Kabsch and Sander[91] which calculates hydrogen bond potentials and correlates them with ϕ- and ψ-angles to match SS elements.

The main difference between the peptides is observed for Tyr^5 which has a higher propensity for turns in the DM compared with the L-DM. The turn is correlated with a higher stability in proteins and could indicate that DM is more stable compared with the L-DM[173, 151]. Table 3.3 shows the end-to-end distance and the radius of gyration(R_g) for all the simulations. The L-DM has a higher probability of being in an extended form compared with the DM. The configurational fluctuations in the

Table 3.3: End-to-End Distance and Radius of Gyration

Residue	L-DM[1]	DM[1]	L-DM[2]	DM[2]
d_{min}	3.68	3.50	3.95	3.83
d_{median}	14.29	12.32	15.59	15.17
d_{max}	21.43	21.24	21.46	21.01
$R_{g(min)}$	4.77	4.84	4.88	5.03
$R_{g(median)}$	6.58	6.53	6.75	6.99
$R_{g(max)}$	8.54	8.62	8.65	8.63

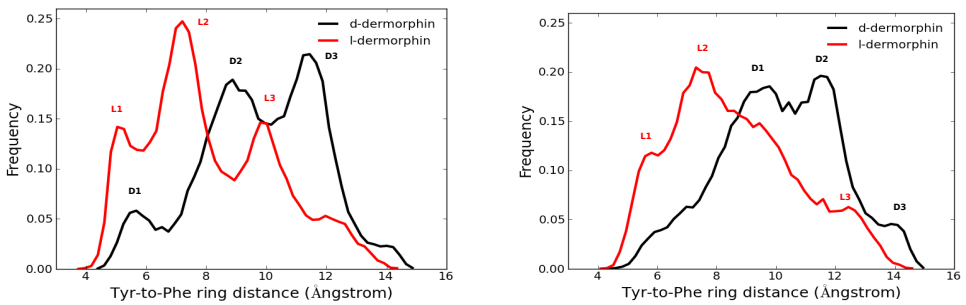
[1] Explicit Molecular Dynamics
[2] Replica Exchange Molecular Dynamics

explicit simulation are higher compared with REMD, indicating a non converged simulation for the same isomers. The R_g varies more between the different simulations and illustrates a discrepancy between the simulations where there is a large variation between the two simulation techniques. In the explicit simulation R_g has smaller values compared with the implicit simulation.

3.4.2 Dynamics of Aromatic Residues in the N-terminal

Table 3.4 shows the side chain torsional χ_1-angles for Tyr[1] and Phe[3]. The χ_1-angle, N-C_α-C_β-C_γ, is analyzed for the two aromatic residues in the N-terminal of the peptides and has three conformations: trans(-180°) or 180°)(t), gauche(-)(-60 °)(g-), or gauche(+)(60 °)(g+). For Tyr[1] there is a trend in the difference between the L-DM and DM where *t* is most favoured. The g(-) is the statistically most preferred χ_1-angle for Tyr in the Dunbrack library[43]. For Phe[3] the conformations are more distributed between *t* and g(-). The *t*-angle is decreased, while g(-) is increased. The g(+) seems to be sampled with the same relative frequency in the four simulations. The g(-) is the statistically prefer-ed χ_1 for Phe in the Dunbrack library, but this could be different for small peptides.

Figure 3-1 shows the distance between the two aromatic rings, Tyr[1]-Phe[3], which is measured as the distance between the center of the two aromatic rings. In L-DM, it is possible to get a shorter distance between the two rings while the D-isomer has a larger distance. This is due to the steric repulsion of the D-ala in the second position.



(a) MD (b) REMD

Figure 3-1: The distance between the aromatic side chains of Tyr[1] and Phe[3]. The distributions are different between the two peptides where the L-DM has its main density between 4 and 6 while the DM has its main density between 8 and 12 .

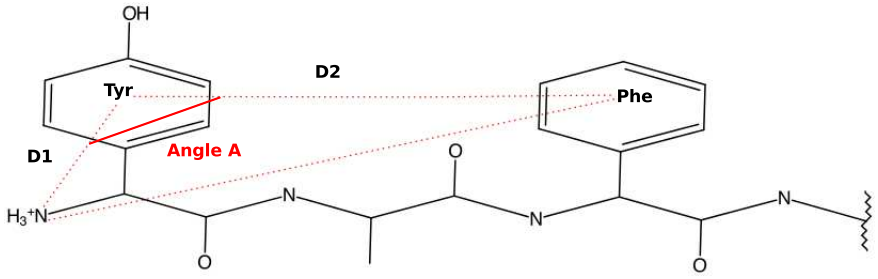
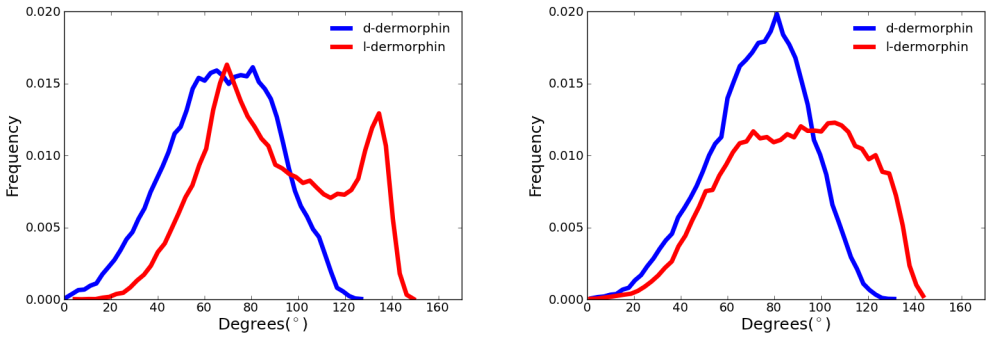


Figure 3-2: The Definition of Angle A. The angle defined by the midpoint of the aromatic residues and the N-terminal nitrogen.



(a) MD (b) REMD

Figure 3-3: Distribution of Angle A. The density of the angle A is confined for DM compared with the L-DM.

Table 3.4: χ_1-angles for Tyr[1] and Phe[3]

χ_1-Tyr[1]	L-DM[1]	DM[1]	L-DM[2]	DM[2]
trans	0.60	0.71	0.61	0.69
gauche(-)	0.21	0.15	0.20	0.13
gauche(+)	0.06	0.04	0.05	0.04
χ_1-Phe[3]				
trans	0.37	0.29	0.43	0.37
gauche(-)	0.37	0.43	0.27	0.35
gauche(+)	0.13	0.16	0.16	0.15

[1] Explicit Molecular Dynamics
[2] Replica Exchange Molecular Dynamics

Figure 3-3 shows the frequency distributions of the angle, A for both the MD and REMD which measure the variability between the two ring systems(see Figure 3-2 for an explanation of angle A). The conformational flexibility is smaller for DM while L-DM has a larger variance for the angle.

Frequency Calculation of Two End-to-End Configurations

Vibrational experiments are being conducted with the two peptides. To complement and assess the experimental work the Hessians are calculated for two different subpopulations of DM from the MD simulations. Table 3.5 contains structural data for the two peptides simulated and Figure 3-4 shows the calculated spectra. The main analysis is on the most intense peaks from the computations as well as the amide I, II, and III bands which indicates SS elements. To assign the frequencies,

Table 3.5: Frequency Assignment for Amide I, II, and III

End-to-End()	Amide I[1]	Amide II	Amide III
Short	1730,1696 cm^{-1}(1666)	1574 cm^{-1}(1516)	1316 cm^{-1}(1267))
Long	1788,1759,1735 cm^{-1}(1671)	1547 cm^{-1}(1490)	1342 cm^{-1}(1292)

[1] Values in Parenthesis are scaled by 0.961

they are scaled by 0.961 cm^{-1} which is an empirical factor for the B3LYP functional for gas phase calculations[206].

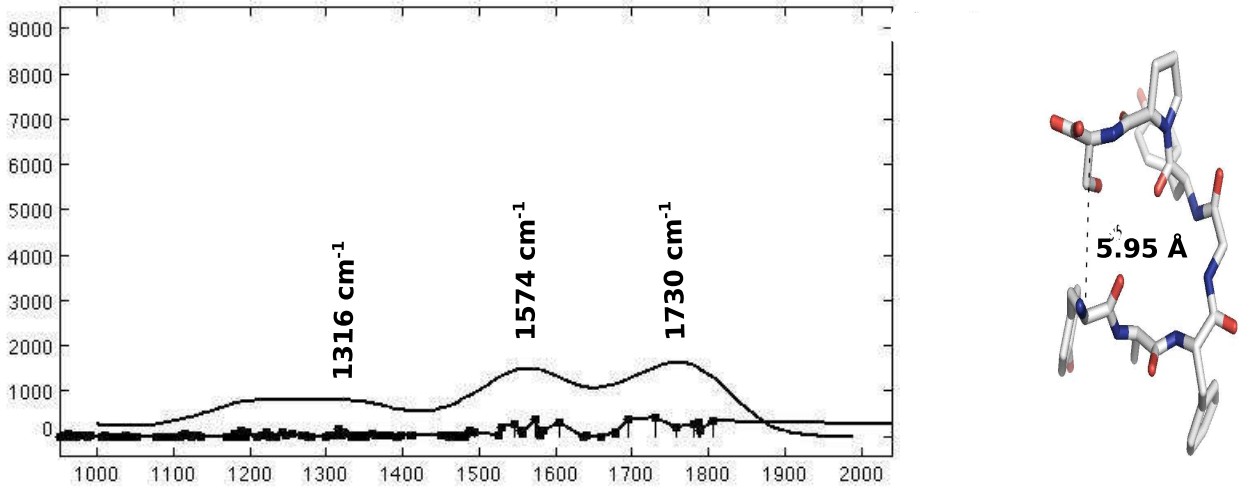
Figure 3-4(b) shows the vibrational frequencies for an extended form of DM with an end-to-end distance of 20.15 Å. 1796 cm^{-1} is composed of contributions made form the N- and C-terminal. The amide I vibration is assigned to 1735 cm^{-1} and is a C=O stretch of the backbone amide carbonyls. 1309 cm^{-1} is a composition of C-H bend from Tyr^1 and Phe^3. 1191 cm^{-1} is a composition of O-H from the Tyr^5 and NH_3^+. The vibrational modes for the short end-to-end(5.95 Å) are shown in Figure 3-4(a) and from the amide I vibration, two main components are identified on the backbone, 1730 and 1696 cm^{-1} which can be ascribed to a β-sheet like structure.

3.5 Discussion

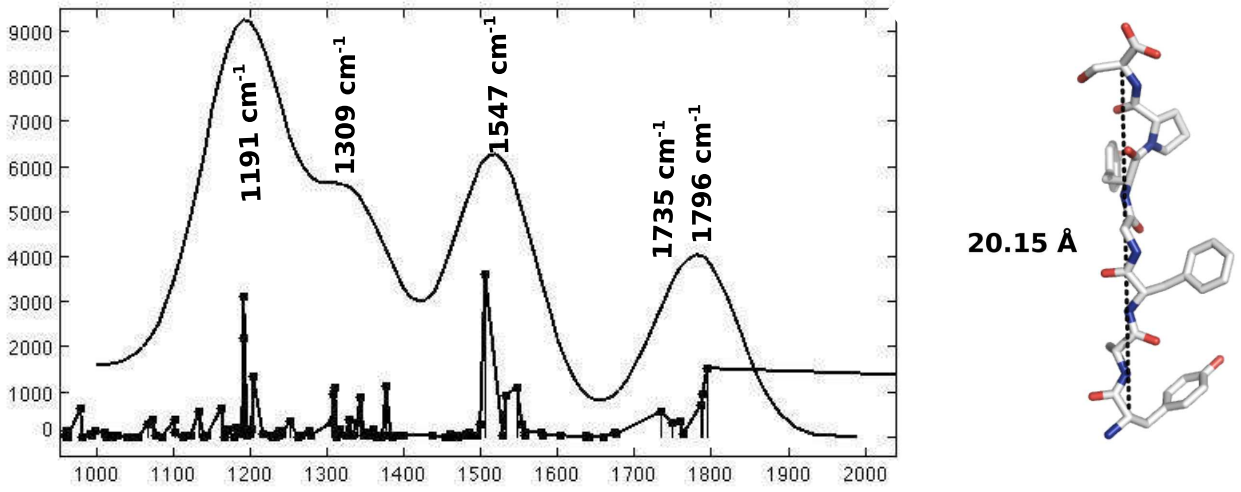
3.5.1 Configurational Difference between L- and D-isomer

The simulations show a difference between DM and L-DM and especially in the N-terminal region it is possible to distinguish between them. Experimentally, it has been shown that the peptide is an ensemble of substructures which are able to move between different conformations[50]. Even though DM has been considered relatively rigid compared with other heptapeptides, e.g. deltorphin, it has many subpopulations. Thus, it has been hard to determine the most probable structure using NMR[33]. The problem in the NMR experiments have not only been the many subpopulations, but also that many of them are linear and hence do not give enough information.

The differences observed are not as significant as would have expected. The vari-



(a)



(b)

Figure 3-4: Frequency Calculations. The amide I, II, and III bands are investigated for three different conformation of the D-isomer. (a) The short is assigned a turn 1736 cm^{-1}(1666) from the amide I vibration. (b) The extended structure is assigned to a sheet with its amide I vibration of 1735 cm^{-1}(1671). The vibrational spectra are plotted using gaussian function with a line width of 50 cm^{-1} in TmoleX[184]

ations have been mostly observed in the N-terminal part where the orientation of the aromatic side chains and their interaction with the charged ammonium group of the N-terminal part show differences. The overall differences in the peptide structure have not been as significant as expected, i.e. the differences in populations of the turn elements are relatively small. The analysis of the secondary elements, end-to-end distance, and the radius of gyration have not shown a significant difference between the two isomers. The perturbation of the polymer seems to be largest around the insertion of the D-enantiomer Ala as seen in the data analysing its influence on Tyr^1 and Phe^3. Here, it is possible to observe a difference between the two peptides where the L-DM is more flexible than DM.

3.5.2 Improvements of Classical Results

The overall volume of phase space has been sampled by both methods, but the probability density has not in all cases converged to the same distribution. The data analysis of the explicit simulations indicate that the simulation has not converged and that the energy landscape is rough with many minima due to the increased degrees of freedom. The measured values from the REMD simulations seem to have converged from block averages made from the trajectories.

The implicit solvation model did not seem to bias the simulation, but it might have underestimated the hydrophobic interactions which have been observed in some simulation with implicit solvent[209]. The short end-to-end distance is not sampled as much as in the explicit simulation and the radius of gyration is also bigger. Experiments indicate that dermorphin is mostly a linear polymer and show with confidence that the extended conformations is the most sampled.

In the explicit model, there is a high frequency of the short polymer and this could be due to kinetic traps imposed by the solvent. This is seen in other simulations with explicit solvent[116].

The parameters of the dihedral angles have been shown to often bias the simulations towards different structures. Here the ff99SB is used which has shown good correlation with experiment, and its parameters are stable for long simulations[171, 120].

Small perturbations can change the structural properties of peptides as shown. Here, MD simulation is a good supplement to both vibrational[105] as well as NMR[109] experiments where it can be hard to determine the relative concentrations of the various species and conformers in the ensembles.

3.5.3 Vibrational Spectroscopy Calculations of D-isomer

From the simulated data, it has been seen that the extended structure is observed for the L-DM and should give rise to vibrational modes in the Amide I band which should be different from those of the DM. Preliminary experimental data indicates that the observations from the MD simulations are in agreement with vibrational experiments.

The calculated vibrational spectra of the three conformations(short, medium, and long end-to-end) of the DM have different vibrational properties. In the case of vibrational spectroscopy, the dipole strength in IR is overestimated as no screening of any of the charges is done in the gas phase calculations.

In the extended polymer, intense vibrations are observed due to strong polarization near the charges at the N-terminal of the peptide. Here, the program Turbomole has been used to compute the spectra which includes the solvation model COSMO for geometry optimizations but is not implemented for Hessian calculations. Including a solvation model might improve intensities.

To interpret the vibrational spectra it is necessary to have an idea of the composition of the observed modes and their relative intensity. Proteins and peptides are big systems and the configurational space is large with many conformations which will give different vibrational spectra.

The techniques are great for small molecules where it is possible to determine all vibrations since there are a limited number of conformations for the substance to populate. For peptides, the size of configurational space gives inherent problems to vibrational spectroscopy. Furthermore, the peptides are in an equilibrium among different conformations and the observed vibrations are a superposition of modes of different structures. The calculations of the Hessian for vibrational spectra along with

the polarization tensor are computationally intensive and puts a limit to how many structures can be calculated.

3.6 Conclusions and Outlook

A difference between the two isomers has been observed and was most pronounced in the N-terminal part of the peptide.
The folding of peptides in different isomeric states has been tried without full convergence. The implicit calculation is close to being converged, while the energy surface for the explicit calculation is more rough and has not converged. The calculations should be extended to reach convergence.
QM calculations have been performed to determine different vibrations and have been assigned accordingly. The calculated spectra will be compared with experimental data from infrared and Raman spectroscopy.

3.7 Acknowledgements

A manuscript has been prepared on the classical data(see Appendix for the manuscript).
The research was design by my supervisor Henrik Bohr(HB, Physics, DTU) and performed by Per Jr. Greisen(PG, Physics, DTU). The data were analysed by HB, PG and Bjorn Nielsen(BN, Physics, DTU). The second part is an evaluation of the simulated data using vibrational spectroscopy: IR and Raman. Experiments were performed by Karin Stibius Jensen(KS,Physics, DTU) and Rolf W. Berg(RB, Chemistry, DTU). The calculated spectra was conducted by PG and the results were analysed by PG, HB, KS, and RB.

Chapter 4

Nitrite Reductase

4.1 Introduction

Nitrogen is vital for life processes and is present in both DNA and proteins. One way to reduce the atmospheric nitrogen, N_2, is through lightning or bacterias which can fixate the nitrogen.

The fixation of nitrogen is part of the nitrogen cycle where the N_2 molecule is reduced to ammonia, NH_3. The cycle can be divided into two parts: nitrogen fixation and denitrification.

Artificial nitrogen fixation is used to generate fertilizer for agriculture and the industrial fixation process is conducted under high pressure and temperature whereas enzymes lower the activation energy for the reduction at 300 K and 1 bar.

One molecule in the nitrogen cycle is nitrite, NO_2^-, which can either be oxidized to nitrate, NO_3^-, or reduced to nitrogen monoxide, NO. Nitrite is shown to be lethal to humans in large amounts and there is a fundamental interest to understand the enzymes involved in the process of reducing or oxidizing the molecule[4]. Two kinds of enzymes are known to reduce nitrite to nitrogen monoxide(NO), cytochrome cd_1 which uses heme for the reduction and the blue-green copper enzymes which use Cu for the reduction. The copper enzymes are the focus of the following study.

4.2 Blue Copper Proteins

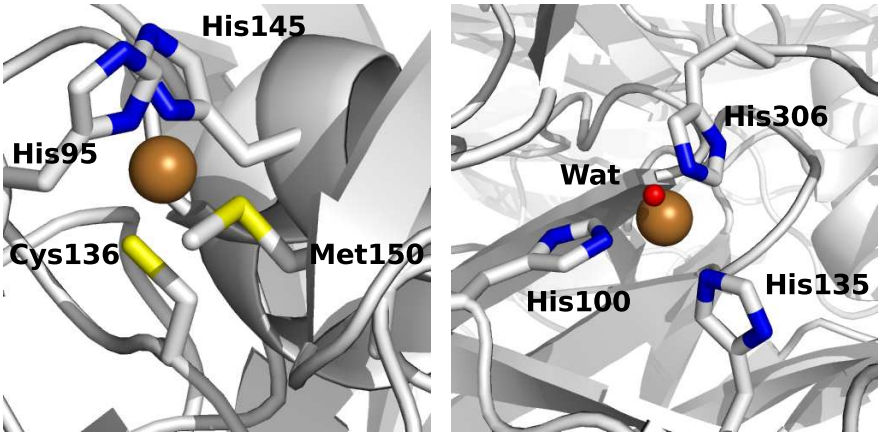
The family of blue copper proteins(BCP) is large and a wide range of different functions is performed. Their name is due to the optical color properties which are related to members of the family having a characteristic blue color though some are green[80]. Their general function is long-range electron transfer(ET) in biological systems. They can be divided into the two groups: mono- and divalent copper-proteins but they have properties in common such as the Type I site[163]. Among the copper BPCs with one Cu are azurin or plastocyanin whereas Nitrite Reductase contains two copper ions. The blue copper proteins have high redox potentials ranging from 184-800 mV which is higher compared with the aqueous Cu(I)/Cu(II) couple having a redox potential of 160 mV[155, 149].

4.3 Nitrite Reductase

The copper containing nitrite reductase is part of the respiratory chain and accepts an electron from cytochrome c or azurin[63, 104]. The enzymes have different colors depending on the bonding characters for one of the copper ions.
The enzyme from *Achromobacter cycloclastes* is a green colored protein and is a 110 kDa homotrimeric protein. Each monomer consists of an eight-stranded beta-barrel topology which generates a hydrophobic surface for the interaction of the monomers. The trimer is held together by strands pairing up with strands from the other monomer. Furthermore, one of the copper ions is coordinated by a histidine 306(His306) from the other monomer. Each monomer contains two copper sites, type I(T1) and type II(T2), which are located $\approx$ 13 Å from each other.

4.3.1 Type I Center

The T1 center is found in all blue copper proteins and is involved in either intra- or intermolecular ET. The Cu is coordinated by four residues, methionine 150(Met150), 2 histidines 95 and 145(His96 and His145) and one cysteine 136(Cys136)(see Figure



(a) (b)

Figure 4-1: The Type I and Type II Center in Nitrite Reductase. In Nitrite Reductase from *Achromobacter cycloclastes* both metal centers are tetrahedral but the Type I site is distorted. (a) The Met150 comes from a β-sheet which could decrease its configurational entropy. In other BCP, the methionine is located on a loop which would make it more flexible. (b) The tetrahedral geometry of the T2 site where the water is replaced by nitrite during the reaction.(PDB ID: 1NIC).

4-1(a)). In other BCP, a glutamine or a carbonyl from the backbone have been observed in the coordination of the Cu-ion.

The site is screened by the protein matrix which protects it from the surrounding water and tunes its redox potential. In NiR, the T1 site is a distorted tetrahedron where the molecular bonding between Met150 and Cu has more σ-character compared with the π-bonding character found in plastocyanin[181, 2]. The geometry of the site is a flattened tetrahedral distorted to a square planar the so-called Jahn-Teller effect. The copper sulfur distance from the coordinating methionine displays a short bond of 2.65 Å compared with the 2.82 Å bond in plastocyanin(PDB ID: 1PLC). This bond is correlated with the bond between Cys136 and Cu which are 2.18 Å in NiR and 2.07 Å in plastocyanin[107, 62]. This bond length difference is able to tune the redox potentials with approximately 70 mV for the different T1 sites[56, 139].

4.3.2 Type II Center

In the T2 site, the Cu is coordinated by 3 histidines 100, 135, and 306 where one of the histidines comes from a different monomer. The fourth ligand is either water

Table 4.1: Atoms Included in Pathways I and II

Path I[1]	N_{ϵ}^{135}	C_{δ}^{135}	C_{γ}^{135}	C_{β}^{135}	C_{α}^{135}	C^{135}	N^{136}	C_{α}^{136}	C_{β}^{136}	S^{136}
Path II	N_{ϵ}^{135}	C_{ϵ}^{135}	N_{δ}^{135}	O^{136}	C^{136}	C_{α}^{136}	C_{β}^{136}	S^{136}		

[1] 135: Histidine 135, 136: Cysteine 136

or nitrite. The site is located between the two monomers in a cleft which is located at the end of a 12 Å deep channel[59]. It has been speculated that residues in the channel are involved in the binding and release of the substrate and product. The type II center is the catalytic site for the reduction of nitrite.
In the T2 site two residues are conserved, aspartic acid 98(Asp98) and His255, which participate in the catalytic cycle by donating protons during the reaction cycle.

4.3.3 Tunneling Pathway

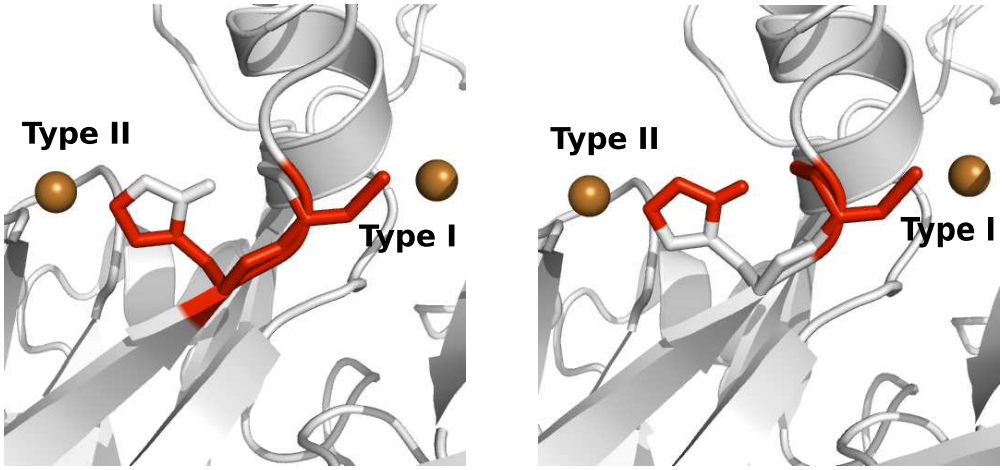
A vital understanding of the reaction mechanism in Nir, is how the electron is transferred from T1 to T2. The electron pathway is believed to go through the backbone of the protein from the T1 to the T2 but other pathways are possible.
Figure 4-2 shows the two most probable pathways for the ET and the atom names and types are given in Table 4.1. Pathway I(P1) is defined as going through the covalent bonds between the two Cu ions whereas Pathway II(P2) goes through a hydrogen bond between the Nδ on His135 and a carbonyl oxygen on Cys136.
It has theoretically been shown that there is a penalty going through a hydrogen bond compared to a covalent bond[18] but depending on fluctuations in the system this penalty can be cancelled out.

4.4 The Reaction Mechanism of Nitrite Reductase

Figure 5-2 illustrates the reaction mechanism and the most important steps. The reaction is simple and consists of one electron and two protons

$$NO_2^- + 2H^+ + e^- \rightarrow NO + H_2O$$

where the chemical reaction is performed at the T2 of NiR. **(Step 1)** In the resting state of the enzyme a water molecule is coordinating as the fourth ligand at T2. The

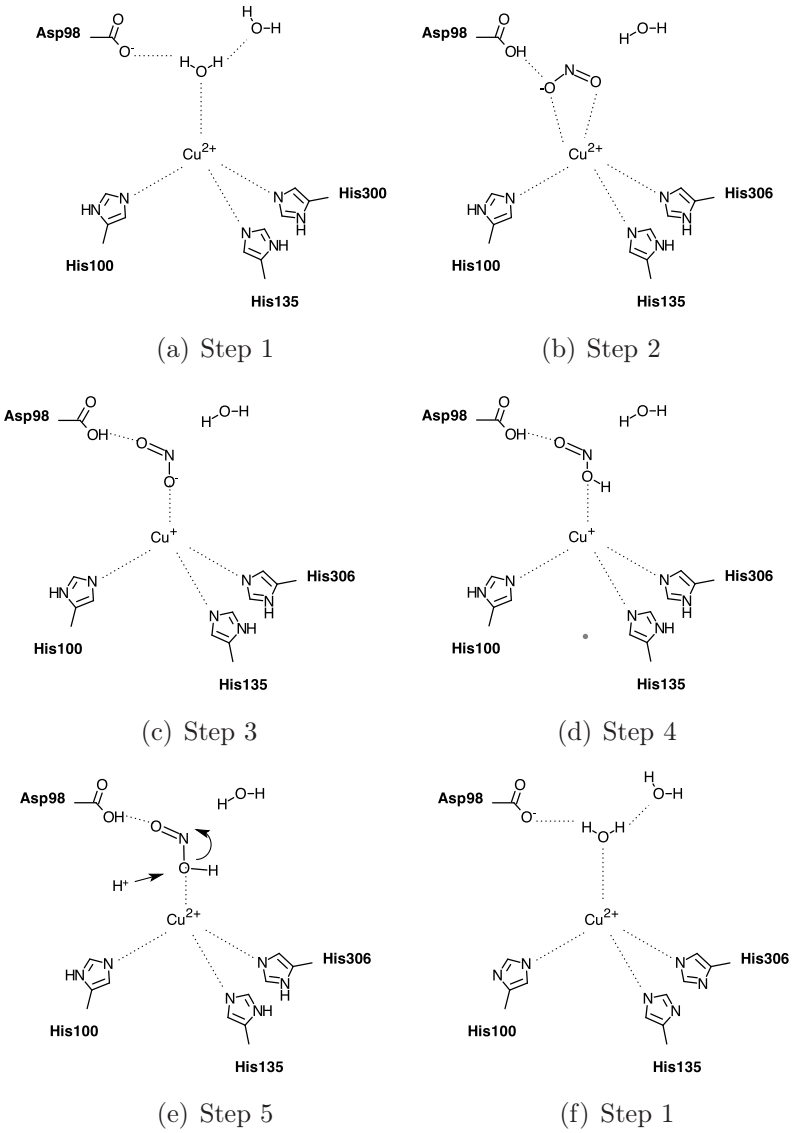


(a) (b)

Figure 4-2: Pathways for Electron Transfer. (a) In pathway I the electron is transferred only through covalent bonds. (b) In pathway II the electron is also transferred through a hydrogen bond between the carbonyl of Cys136 and the Nδ of His135.

water is fixated by a nearby acid, Asp98, which makes a strong hydrogen bond with the water molecule. **(Step 2)** A nitrite molecule replaces the water molecule and the copper ions remains in the oxidized state. The binding mode of nitrite has been discussed but consensus is that it binds in a bidentate fashion to the Cu^{2+}[186]. Infrared spectroscopy studies indicate that Asp98 is deprotonated in the oxidized form[208] and the pKa value of the Asp98 has to be raised. For this to happen, the substrate needs to be protonated or the pKa-value of the Asp98 has to be raised. It has been proposed that nitrite enters the channel, which changes the local charge from +1 to 0. This lowers the redox potential of the T2 and increases the pKa value of Asp98 and it gets protonated[128]. The protonation is confirmed by the experimental substitution of Asp98 with asparagine, which still has activity but is lowered[152].
(Step 3) The T2 is reduced where an electron is transferred from the T1. The ET is a spontaneous process as the redox potential of the T2 is higher than T1. The rate of the ET is in general very fast and for Cys-His bridge in laccase it is $> 10^3 s^{-1}$[31], which is similar to the ET in NiR.
(Step 4) Two protons have been introduced, protonation of Asp98 and the substrate. The protonation of the substrate is believed to be through a water molecule. The last proton could be coming histidine 255 which is located nearby[128]. **(Step 5)**

Nitrite is reduced to "nitrous" and there is a rearrangement of electrons. A proton is donated to the hydroxyl, OH^- giving water. As the overall charge is at the T2, the pKa of Asp98 decreases and the proton is released.



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

(e) Step 5 (f) Step 1

Figure 4-3: Reaction Mechanism of Nitrite Reductase(Figure is based on Marothy et al.[128])

4.5 Aim of Research

The aim of the following research is to investigate the changes induced on the trimeric protein by reducing the T1 or the T2. Quantum mechanical calculations are used

to estimate the charge distribution of a model of the bridge system. The distance between the two Cu ions is analysed and the relative probability for the two ET pathways to occur is estimated.

4.6 Methods

4.6.1 QM Optimization of Bridge Model

A model of the bridge was constructed from the crystal structure(PDB ID: 1NIC) and optimized. The model contained the following side chains: His95, His100, His145, Met150, and His306, which were methylated at the C_β to decrease the number of atoms. A proton was added to the Cys136 residue at the backbone carbonyl while a formyl-group was added to the N-amide of residue His135 to model the strain on the N-H bond. The model consisted of 105 atoms including the two Cu ions and either a water or nitrite molecule.

Two different functionals, BP86[15, 146] and TPSSH[148, 183], were used to optimize the model. In the BP86 calculation, the def-SVP[167] basis set was used, while for TPSSH, the def-TZVP[201] was used which include diffuse functions. To model the electrostatic screening from the protein, the COSMO solvation model[97, 168] was used with a dielectric constant of 10 which has been shown to be reasonable for geometry optimizations[87]. The radius of the Cu atoms was set to 2.0 for screening effects[87].

The overall charge of the model depended on the ligand as well as the redox state of the copper ions. With both coppers in the oxidized state the overall charge of the model was +3 with water and +2 with nitrite. Adding an extra electron to the system decreased it with 1.

The structure was first optimized in the highest possible spin state where both copper atoms were in the d^9-state giving a spin of 1(M_S=1). After convergence the structure was optimized pairing the two electron(M_S=0). In the low spin state model, an extra electron was added and reoptimzed to monitor charge distribution of the

model(M_S=$\frac{1}{2}$). The simulations were performed with Turbomole 5.9[3] and all charges were fitted using the electrostatic potential(esp) Kollman-Merz in Turbomole.

4.6.2 Parameters for Reduced Type II Center

The force constants and electrostatics for the reduced T2 were not available and were calculated for the later MD simulation. The T2 center was isolated from *Achromobacter cycloclastes*(PDB ID: 1NIC) and the histidine residues were truncated at the C_β atom and hydrogens were added. The overall charge was +1 and the Cu was in the d^{10}(M_S=0).

The geometry was optimized using the B3-LYP functional[185, 16, 115, 114] with the double-ζ basis set[166] for Cu which was enhanced with diffuse p, d, and f functions with the exponents 0.174, 0.132, and 0.390. The rest of the atoms were optimized with 6-31G*.

From the geometrized structures, the hessian matrix was computed and the force constants were obtained using the algorithm by Seminario[170] implemented in Hess2FF developed by Nilsson[136]. The electrostatics was obtained from the esp Kollman in Turbomole and the charge of the C_β atoms was adapted so the overall model had the same charge as in the QM calculation(All parameters are collected in Supplementary Material 4.10.2).

4.6.3 Molecular Dynamics Simulation of Nitrite Reductase

The crystal structure of nitrite reductase from *Achromobacter cycloclastes*(PDB ID: 1NIC) was used with a resolution of 1.9 Å. To generate the trimer, the coordinates were transformed according to the symmetry operations available from the pdb-file.

To assign protonation states of the histidine residues, the hydrogen bond network was manually optimized. The following residues were protonated on the Nδ: 26, 100, 135, 231, 255, 306, and 319 where 100, 135, and 306 are metal coordinating residues. The other histidines were protonated on the N_ϵ: 53, 76, 95, 145, 217, 245, and 260 where 95 and 145 are the metal coordinating residues. For four of the residues, different

protonation states were chosen: 28, 217, 245, and 260, compared with Kerpel and Ryde[92](see Supplementary Material 4.10.1).

The parameters developed by Kerpel and Ryde[92] for the reduced and oxidized T1 site were used where the parameters for the reduced site was develop for plastocyanin. For the reduced T2 site the parameters were developed as described earlier.

Three different simulations were set up. In simulation 1(sim 1), all the copper ions were in their oxidized state. In simulation 2(sim 2), the T1 site is reduced in one of the monomers while the rest of the copper ions were in their oxidized state. In simulation 3(sim 3), the T2 site is reduced in one of the monomers keeping the rest oxidized.

Crystal waters overlapping with the protein were removed giving 174 water molecules. The remaining crystal waters were minimized for 10.000 steps with 5.000 SD and CG where the protein was fixed adding a constraint of 500 kcal/mol. Next, the whole complex was minimized 10.000 step with 5.000 of SD and CG. The complex was solvated in a square box adding 15 Å giving a box size of approximately 120x120x120 Å for the three complexes. All water molecules were treated as TIP3P[90]. In sim 1, the complex was neutralized by adding 27 Na^+ ions. In sim 2 and sim 3, 28 ions were added. This gave a total of 148.662 atoms in sim 1, 162.985 atoms in sim 2, and 164.149 in sim 3. All water molecules were minimized for 10.000 steps with 5.000 SD and CG where the protein was fixed adding a constraint of 500 kcal/mol. Next, the whole complex was minimized 10.000 steps with 5.000 of SD and CG. To minimize solvent fluctuations, the system was heated up from 0 to 300 K for 100 ps using an integration step of 1 fs keeping a weak constraint of 10 kcal/mol on the protein. The volume was kept constant to equilibrate the pressure. The system was equilibrated for 1 ns in the NPT ensemble keeping the pressure at 1 atm and the temperature constant at 300 K using the Langevin thermostat with a collision frequency of 1 ps^{-1} and an integration step of 2 fs. All hydrogens were kept fixed using SHAKE[161] and a cut-off of 10 Å was used for all non-bonded interactions. Full electrostatics were applied using the Particle Mesh Ewald and periodic boundary conditions[35]. 1 ns was used as equilibration and the MD simulation was continued for 10 ns using

Amber 10[26] and the ff99SB force field[76].

4.6.4 Electron Pathways

To estimate the probability distribution for the tunneling pathways as well as coupling between T1 and T2, the *Pathways* plugin for VMD was used. The algorithm was an implementation of the model developed by Beratan et al.[17].

Briefly, the algorithm approximated the protein as a graph which was connected either by covalent bonds(CB), hydrogen bond(HB), or a through space(S) term with a penalty assigned to each step.

The penalty was ranked in the following order

$$CB < HB < S \tag{4.1}$$

and the graph was optimized finding the shortest distance between T1 and T2. An estimate of the donor-to-acceptor coupling was computed

$$T_{DA} = \Pi_i \epsilon_i^{CB} \Pi_j \epsilon_j^{HB} \Pi_k \epsilon_k^{S} \tag{4.2}$$

where ϵ was the penalty for the propagation.

The default parameters were used in the computation with $\epsilon_i^{CB} = 0.6$ and the HB having a penalty of

$$\epsilon_j^{HB} = \epsilon_i^{CB} * \exp[-1.7(R_j - 2.8)] \tag{4.3}$$

where R_j was the distance in Å. The through space was defined as

$$\epsilon_j^{S} = \epsilon_i^{CB} * \exp[-1.7(R_j - 1.4)] \tag{4.4}$$

Table 4.2: Charge Distribution of Cu and Ligand Atoms in the QM Model

Atoms	Cu(I)	Cu(II)	O(H_2O)	N(NO_2^-)	Energy(a.u)
WaterA(M_S=1)	0.396714	0.506642	-0.707190		-6519.132306
WaterA(M_S=0)	0.354917	0.424063	-0.703746		-6519.121300
WaterB(M_S=$\frac{1}{2}$)	0.469469	0.255996	-0.710087		-6519.335476
WaterA(M_S=1)TPSSH	0.490784	0.633089	-0.813939		-6522.000698
WaterA(M_S=0) TPSSH	0.481106	0.632562	-0.814591		-6522.000734
WaterB(M_S=$\frac{1}{2}$) TPSSH	0.532681	0.199869	-0.861828		-6522.190839
NitriteA(M_S=1)	0.393039	0.042123		0.615453	-6648.018709
NitriteA(M_S=0)	0.399774	0.067867		0.611759	-6648.005266
NitriteB(M_S=$\frac{1}{2}$)	0.445446	-0.163052		0.628724	-6648.190270
NitriteA(M_S=1)TPSSH	0.476664	0.158399		0.650948	-6651.002079
NitriteA(M_S=0) TPSSH	0.444394	0.117460		0.647916	-6650.964736
NitriteB(M_S=$\frac{1}{2}$) TPSSH	0.482854	-0.040774		0.685713	-6651.168032

4.7 Results

4.7.1 Charge Distribution of the Type I and Type II Centers

Table 4.2 shows the charge distribution from the QM calculations on the bridge model atoms. Both models have lower electronic charge on the Cu at the T2. Adding an electron to the system, the electronic density is localized on the Cu ion at the T2 state.

The coordinating water molecule is relatively unaffected by the extra electron in the model with 0.03 in the high spin state model compared with 0.07 e in the low spin state model using the BP86 functional. There is a qualitative agreement between the BP86 and the TPSSH functional although the values differ. Replacing water with nitrite lowers charge of the Cu at the T2 center and the charge of the Cu becomes negative as an extra electron is added to the model. The charge distribution of the atoms connecting the two Cu-ions is investigated and the charge of the S^{136}(Cys136) becomes negative.

Table 4.3: Distances in Redox States

	Atoms	Cu-Cu (Å)	S_{M141}-$N\epsilon_{H145}$(Å)
1NIC	$Cu(I)^{2+}$-$Cu(II)^{2+}$	12.72	3.46
Sim 1	$Cu(I)^{2+}$-$Cu(II)^{2+}$	12.71	3.83
Sim 2	$Cu(I)^{2+}$-$Cu(II)^{2+}$	12.74	4.05
Sim 2	$Cu(I)^{+}$-$Cu(II)^{2+}$	12.65	3.60
Sim 3	$Cu(I)^{2+}$-$Cu(II)^{2+}$	12.71	4.28
Sim 3	$Cu(I)^{2+}$-$Cu(II)^{+}$	12.73	4.78

4.7.2 Molecular Dynamics Simulation of Trimeric Nitrite Reductase

The ET rate between the metal centers is exponential proportional to the Cu-Cu edge-to-edge distance[174].

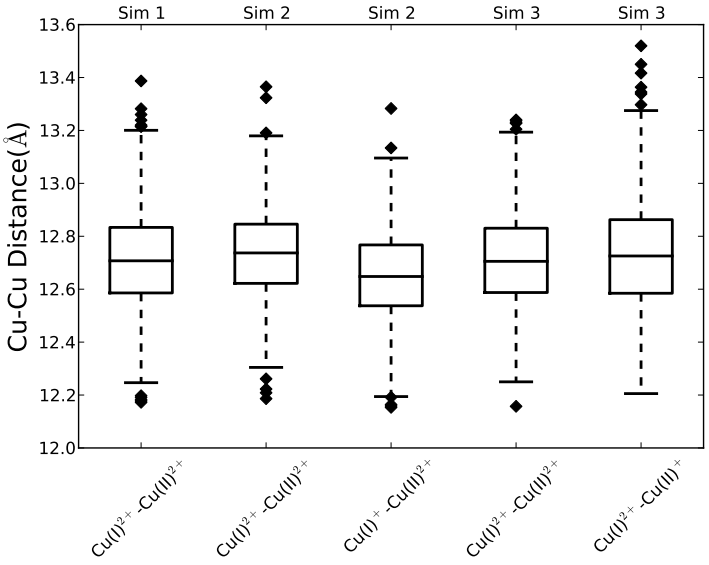


Figure 4-4: The Cu-Cu distance. The distance is illustrated by a boxplot. In sim 2 with the reduced T1 site and the oxidized T2 site, $Cu(I)^{+}$-$Cu(I)^{2+}$, the median is shorter compared with the other redox states.

Figure 4-4 shows a boxplot of the distance distributions and Table 4.3 summarizes the median of the Cu-Cu distance. In sim 1, the equilibrium value of the Cu-Cu distance is 12.71 Å which is close to the value measured in the crystal structure 12.72 Å. As the T1 center is reduced, the distance between the Cu ions start to differ in the

trimeric complex. In the reduced monomer the distance becomes smaller, 12.65 Å, whereas it increases for the other Cu-Cu distances in the trimer, 12.74 Å. In Sim 3, the distance between the two Cu ions is longer in the monomer where T2 is reduced while the distance for the oxidized sites is similar to Sim 1.

To investigate the influence of the fluctuations on the electron tunneling, the pathways between Cu are determined. Table 4.4 shows the probabilities for the different pathways in the trimeric protein complex. The preference between the two pathways is approximately 0.5 with fluctuations in the different simulations. Reducing the T1 site changes the preference between the pathways where P2 is more favoured compared with P1. Having the T2 site reduced switches the preference between the pathways and the probabilities are similar to the oxidized monomers.

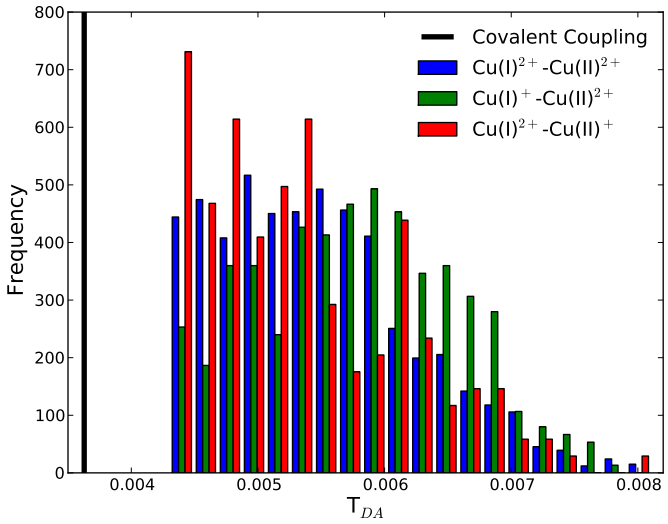


Figure 4-5: The Coupling Strength, T_{DA}, between Donor and Acceptor. Pathway I has the largest electronic coupling. For pathway II the coupling fluctuates more and differs among the different redox states of the protein.

Figure 4-5 shows the distribution of the coupling between the different sites and their redox state. The coupling strength between donor and acceptor is largest for the ET through covalent bonds which corresponds to P1. Interestingly, there is a large coupling in PT2 for the reduced T2 which would suggest that reverse ET could take place which has been observed[49].

Table 4.4: The Probabilities for the two Pathways in Nitrite Reductase

	Redox States	Pathway I	Pathway II
Sim 1	$Cu(I)^{2+}$-$Cu(II)^{2+}$	0.57	0.43
Sim 2	$Cu(I)^{2+}$-$Cu(II)^{2+}$	0.45	0.55
Sim 2	$Cu(I)^{+}$-$Cu(II)^{2+}$	0.20	0.80
Sim 3	$Cu(I)^{2+}$-$Cu(II)^{2+}$	0.45	0.55
Sim 3	$Cu(I)^{2+}$-$Cu(II)^{+}$	0.64	0.36

4.7.3 Radial Distribution Function of the Type I Center

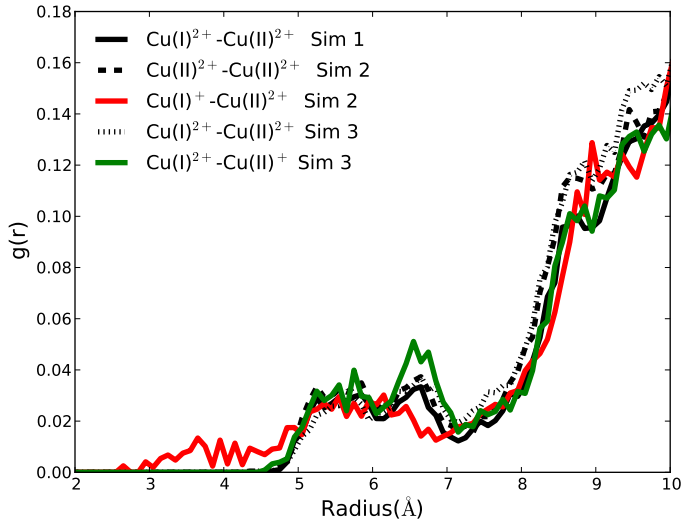


Figure 4-6: The radial distribution of Cu-ion in T1 and Water. In the reduced state of T1, it is possible for water to approach the copper ion. In the oxidized state of T1, all simulations have the same features for the rdf screening the metal.

Figure 4-6 shows the radial distribution function(rdf) for the 3 simulations. For the oxidized T1, the rdf has the same feature in all of the simulations and starts to increase after 4 Å. By reduction of the T1 site, the rdf has small fluctuations around 3 Å. The space for water molecules is due to fluctuation of the turn, residues 136-138, coming off a β-sheet.

The fixation of the His145 is investigated as its dipole strength has been correlated with the redox potential[25]. Its Nδ is coordinating the metal ion while the Nϵ is stabilized by either a water molecule or a sulfur atom from Met141. The fluctuation of the S_{M141}-Nϵ_{H145} is affected by the oxidation step of the copper ion.

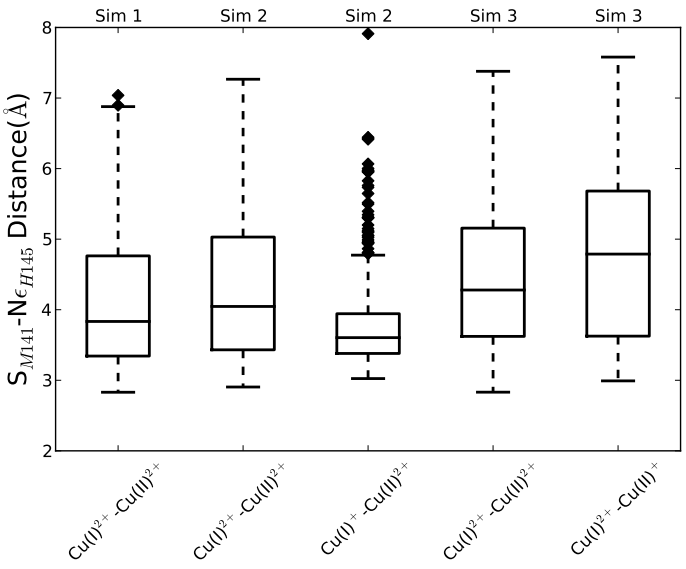


Figure 4-7: The S_{M141}-Nϵ_{H145} Distance. The distances are plotted as a boxplot. In the sim 2 with the reduced T1 site and the oxidized T2 site the distance is lower than observed in any of the other simulations. The first and third quartile are more confined compared with the rest of the simulations.

Figure 4-7 shows fluctuations of the S_{M141}-Nϵ_{H145} distance and the median values are displayed in Table 4.3. The distance fluctuates more compared with the Cu-Cu distance which is more stable. Reduction of the T1 site decreases the S_{M141}-Nϵ_{H145} distance while it is increased by reduction in the T2 site. Waters are observed interacting with the dipole from His145 when T1 is oxidized. The distance between Met150 and the Cu in the T1 site has been investigated experimentally and theoretically and has been discussed thoroughly(see [139, 56]). As mentioned, the bond varies much and it is not clear how much restrain the protein has on the bond. Here, the bond distance is monitored during the simulation in the three systems. The distance fluctuates around the equilibrium value used, 2.64 Å. To test the strain by the protein on the bond, the force constant is removed in the simulation. The bond length is increased to 2.8 Å and fluctuates around this value. No strain is observed from the protein to shorten the bond to the value observed in the crystal structure.

4.8 Discussion

4.8.1 QM Calculations of Nitrite Reductase Models

The model system is cable of capturing the redox properties where an excess electron is mainly localized at the T2 site. It has not been possible to determine the most favourable pathway for the ET from this model. For the bridge atoms, the main charge localization is on the Cys136 sulfur atom while there is a charge decrease on Nϵ of His135.

To capture the ET pathway would require time perturbations as in TD-DFT or computation of diabatic states. The calculation performed here corresponds to an adiabatic system where the electron is mainly delocalized on the potential energy surface. This delocalization is further enhanced by using DFT which inherently delocalizes it. The pathways identified are hard to determine at a much higher level of theory. It has been discussed to calculate the configurations in a diabatic state and then use non-equilibrium Green functions to estimate the rate transfer for the electron. This has been achieved with some success using semi-empirical methods[55]. The problem is how to split up the system into fragments and keep it diabatic. Here, an adiabatic calculation was performed for the system which confirmed the redox properties of the small model system. To calculate the diabatic state of the modelled system multiconfiguration calculations were performed but electron delocalized and no convergence was obtained. The water or nitrite molecule is not affected by the excess electron in the system which is to be expected. The energy barrier needed to reduce the substrate requires fluctuations. Furthermore, important catalytic residues have not been included which are necessary to model the reaction. Protons have to be donated to the substrate to lower the activation energy for ET[128]. The calculation has shown which part of the bridge model is most probable for accepting electrons which agrees with T2 having a lower redox potential than T1[138].

4.8.2 Functionals and Basis Sets used in the Simulation

Two different types of functionals are used, BP86 - a generalized gradient approximation(GGA), and TPSSH, a hybrid meta-GGA functional.

The two functionals are used to get a confidence interval for the calculations where the GGA tends to overbind while hybrid meta-GGA can underestimate binding. The TPSSH functional has shown good correlation with experimental energies for second row transition metal[86].

The basis set errors have not been tested systematically for the model but the overall trend for the model system has been captured as seen by the agreement between the two computations. To use a larger basis set with the BP86 functional is probably beyond the precision of the functional and not necessary. Qualitative agreement with the applied basis set has been observed in other calculations[202], hence there is confidence in the accuracy and precision of the results. The TPSSH functional can give wrong energies for various spin states where the BP86 has been shown to be within theoretical uncertainties[86]. The agreement between the calculations increases the confident in the results.

4.8.3 Molecular Dynamics Simulation of Enzyme

The MD simulations show changes as the redox states are altered in the enzyme. The edge-to-edge distance between the two Cu ions is decreased as the T1 site is reduced. The distance is decreased by 0.3 Å which is a small effect, but small differences can increase the ET rate. It has been shown that the overlap and thereby the electron transfer rate is exponential dependent on the distance where the distance decay is exponential[74, 110, 178].

The electronic coupling between the two sites differs among the redox states of the enzyme. The population of the pathways in the enzyme are very dependent on the redox states. There is a large increase in the population of P2 as the T1 site is reduced. This pathway has a lower electronic coupling for the electron to transfer via hydrogen bond between the backbone carbonyl oxygen of Cys136 to the Nδ hydrogen

of His135. The reverse transfer from the T2 to the T1 has been observed but due to change in the driving force of the reaction[48, 49]. The pathway model is simple in many of its parameters but it has shown to be a very useful method for estimating long-range ET[18].

The His145-Met141 interaction is influenced by the redox state of the T1 center in the enzyme. The fluctuations decrease and the dipole-interaction becomes stronger. This is seen in accordance with other simulations of the T1 site where a stronger polarization of this bond is related to the redox potential of the Cu ion[25]. The bond in the crystal structure is 3.46 Å which is shorter than seen for any of the corresponding median distances in the simulations. The B-factor for this residue is significantly higher compared with the rest of the residues around the T1 and in other crystal structures the bond is 4.06 Å(PDB ID: 2DWT)[82].

Reduction of the T1 site changes the rdf of the Cu-OH_2 compared with the other oxidation states of the protein. There are fluctuations of rdf which increase around 3 Å while the oxidized state of T1 is screened. The smaller screening of the T1 site can affect its ET rate by increasing reorganization energy which has been seen when the T1 site became more accessible to water[203].

4.8.4 Metal Parameters and Time Scale used in the MD Simulation

To simulate transition metals using a classical approach like MD is problematic as many of the metal properties are hard to capture classically, e.g. the symmetry distortions in the copper sites. With the method used here, it is not possible to have dynamical bonds which could switch between the different distortions of the metal site. Especially, the two sulfur bonds in the T1 site have potential to elongate and shorten but in this simulation a single equilibrium distance is chosen.

The parameters estimated by Kerpel and Ryde[92] have been used. In the case of the oxidized T1 and T2 states all parameters have been developed using PDB ID: 1NIC. In the simulation by Kerpel and Ryde[92], an equilibrium distance of 2.42 Å

has been used between Met150 and Cu which is the equilibrium distance from the QM optimization. Here is used the equilibrium value from the crystal structure, 2.64 Å with the same force constant for the bond as used in the simulation by Kerpel and Ryde[92]. This decreases the configurational flexibility of the Met150 but keeps the simulation around the equilibrium value for the crystal structure. The 2.64 Å has been chosen as the equilibrium length so as not to perturb the protein structure by decreasing the bond length which would affect the reorganization of the site.
Test computations have been performed keeping no restraints on the bond and it has lengthened to 2.8 Å.
For the reduced T1 site, the parameters from plastocyanin have been used. The difference between the two proteins is significant in the oxidized state as described. In the oxidized state, there is a Jahn-Teller distortion for the T1 site in Nir while the T1 site is trigonal for plastocyanin[92]. In the reduced state, the differences are smaller as the Met150 bond weakens in Nir as well and even in the oxidized state the bond is relatively weak[179]. Generally the T1 site in Nir is less entatic compared with plastocyanin[107] and it is crude approximation.
As mentioned earlier the data were generated from 10 ns of simulation which are to be extended in future work to get a better statistics. The trimeric complex is used which adds to the statistics but longer trajectories are necessary. The issue of ergodicity is different compared to the folding of peptides.

The starting structure is initially close to the crystal structure(at least for the oxidized state) and during the MD simulation it continues to be close to the structure. Experiments have shown that the reorganization energy is small[203] and gives confidence in that the different redox states simulated should be structurally close to the starting structure.

4.8.5 Outlook

These preliminary data have shown very interesting results. The step is to determine if the observed data are consistent with all available spectroscopic data. The MD simulations will need to be extended to 2 x 50 ns to validate the observed differences

seen.

Furthermore, classical parameters have been developed for the T1 site with nitrite as the fourth ligand in both the oxidized and reduced state. It could be interesting to monitor the dynamical changes of the enzyme by replacing the water with nitrite.

To quantify the two time constants for the two pathways one could try to simulate the excess electron model using TD-DFT.

4.8.6 Conclusions

The QM computations show the charge distribution of the two centers on an adiabatic surface. It is shown that the T2 site has a larger localization of charge when an electron is added to the model.

MD simulations have shown that reduction of the T1 site affects the average distance between the two Cu ions making their distance shorter. Two pathways for the ET have been identified and one has observed that the equilibrium between the pathways are affect by the redox states in T1 and T2.

4.9 Acknowledgements

The research was performed by PG under the supervision of Professor Jens Ulstrup. Great comments on the parametrization of the reduced T2 site was made by Professor Ulf Ryde, Lund University.

4.10 Supplementary Data

4.10.1 Protonation of Histidines

Different protonation states were chosen for the following residues: 28, 217, 245, and 260 compared with Kerpel and Ryde[92]. In their simulation, an extended monomer was simulated and they protonated the residues on both sites which might be reasonable in that case but not in this case.

For His28, the χ_1-angle was rotated from 90° to -90° which optimized the hydrogen bonding network by forming a hydrogen bond from the N_δ hydrogen to the backbone carbonyl oxygen of His26. In addition, the N_ϵ accepts a hydrogen bond from the hydroxyl-group hydrogen of Tyr72.

His217 was protonated at N_ϵ which allows the formation of hydrogen bond to the OD2 atom of Asp53. It accepts a hydrogen bond to its N_δ from the hydroxyl group of Thr216. His245 was protonated on N_ϵ donating a hydrogen bond to OE2 on Glu180 while accepting a hydrogen bond from hydroxyl group of Tyr303. His260 was protonated on N_ϵ donating a hydrogen bond to the carbonyl oxygen of His100 while accepting a hydrogen bond from a water molecule(residue number 1293).

4.10.2 Parameters for the Type II Site with H_2O

Table 4.5 contains the bond and angle equilibrium values obtained from the QM simulation while the fitted charges are collected in Table 4.6. The approach is described in 4.6.2.

Table 4.5: Equilibrium Parameters for Type 2

Bonds	(pm)
His N_ϵ - Cu	203
O_{H_2O} - Cu	258
Bonds	**Force Constants(kcal/mol)**
His N_ϵ - Cu	42.49
O_{H_2O} - Cu	2.31
Angle	**θ(Degrees)**
His C_δ-N_ϵ - Cu	128.46
His C_ϵ-N_ϵ - Cu	125.54
His N_ϵ - Cu -His N_ϵ	119.33
His N_ϵ - Cu - O_{H_2O}	94.60
H_{H_2O} - O_{H_2O} - Cu	101.06
Angle	**Force Constants(kcal/mol)**
His C_δ-N_ϵ - Cu	23.30
His C_ϵ-N_ϵ - Cu	23.51
His N_ϵ - Cu -His N_ϵ	5.72
His N_ϵ - Cu - O_{H_2O}	7.44
H_{H_2O} - O_{H_2O} - Cu	4.70

Table 4.6: Charges used in the Simulation for the Reduced Type 2

Atom	His	H_2O	Cu^+
N	-0.416	O: -0.761	0.079
H	0.272	H: 0.406	
C_α	-0.058		
H_α	0.136		
C_β	-0.352		
H_β	0.173		
C_γ	0.356		
N_δ	-0.353		
H_δ	0.370		
$C_\epsilon 1$	-0.027		
$H_\epsilon 1$	0.177		
$N_\delta 2$	-0.062		
$C_\delta 2$	-0.314		
$H_\delta 2$	0.186		
C	0.597		
O	-0.568		
Total	0.290	0.051	0.079

Chapter 5

Enzyme Design

5.1 Introduction

Enzymes are great catalysts and can accelerate reaction rates by many order of magnitude[84]. The potential use of enzymes is enormous and even today they are used in many processes ranging from medical treatment, washing powder to industrial processes. People are working to extend their uses and replace surface catalysts with biocatalyst which are sustainable and cheaper to use.

The problem is that enzymes do not exist for all synthetic reactions and need to be created. Different laboratory methods have been used to optimize or develop new enzymes[9, 81] e.g. directed evolution(DE). To rationally design new biocatalysts has tremendous potential. This requires that one is able to correlate the protein sequence and function to find the most likely places for optimization. One way to rationalize the design process is to use computer algorithms. The use of protein design algorithms has been extensively developed and used, but only recently have there been success stories[159, 89, 176]. In these studies, three classical chemical reactions: Retro-Aldol, Kemp eliminase, and Diels-Alder, have been catalysed using de novo designed enzymes.

In general, all the de novo designed enzymes have shown low catalytic rates and it has been necessary to optimize them using experimental methods[94, 23].

5.1.1 Zn-Containing Enzymes

Metals have played a big role in the development of industrial catalysts and are still used today[28].

Metalloenzymes use metal ions in catalysis to lower the reaction barrier or to obtain the right redox-potentials. One of the most used metal ions in biology is the zinc ion(Zn) which is used in a variety of ways like catalysis, stabilization of proteins, or as a neurotransmitter. Zn is coordinated by the protein in mainly two different point groups, tetrahedral(T_d), and trigonal bipyramidal(T_{BP}). The reactivity of the site is regulated by the coordinating groups of the metal which are normally histidines(His), aspartic acids(Asp), glutamic acids(Glu), or cysteines(Cys) . The coordination sphere of the metal ion is flexible and through the different ligands, it is possible to tune the reactivity of the site[192]. In enzymes Zn acts as an oxyanion hole like in alcohol dehydrogenase[193], or it activates a water molecule to perform a nucleophilic attack on a substrate like in carbonic anhydrase[30].

5.1.2 Phosphotriesters

Only a short description of phosphotriesters(PT) is given as they will be described more deeply in the next chapter(see Chapter 6). Briefly, it is a phosphate atom with the three ethers bound to it and a phosphoryl group making it a triester.

PTs are used in many industrial processes, as pesticides, or even as chemical warfare agents. The PTs are decomposed through hydrolysis where a hydroxyl anion(hydroxyl) attacks the phosphate atom. The products after the hydrolysis is an alcohol and a phosphodiester.

Due to the industrial applications of PTs, it is not possible to prohibit its production and it is necessary to be able to decompose PTs into non harmful substances.

PTs are easy to develop and can be used as a chemical weapon by terrorist groups which has been done in Tokyo. Hence, there is a great need to be able to degrade PTs efficiently in a safe and reliable way.

Aim of Research

The aim of this research is to computational repurpose[1] mononuclear zinc sites into having a new function by resculpting the protein scaffold.

5.2 Methods

Figure 5-1 summarizes the protocol used to computationally repurpose enzymes.

5.2.1 Selection of Protein Scaffolds

To select protein scaffolds suitable for design, all protein structures from the Protein Data Bank(PDB) were collected. The search was limited to high resolution structures(< 2.9 Å) and a sequence identity < 90 %. The proteins should be expressed in E. Coli and contain Zn in the structures.

A filter was developed to removed non-catalytic sites as well as surface bound zinc ions or multiple zinc sites. All Zn sites having less than 3 atoms from the protein were removed. Structural sites defined as being coordinated to 4 cysteine residues were removed. The PDB files were modified by removing redundant protein chains. Alternate side-chain positions or atomic coordinates were discarded keeping one protein chain along with its metal site. The total number of collected proteins was 105 and included a great variety of protein folds and enzymes classes. All filters were implemented in Python.

5.2.2 Positioning of the Transition State Model

Models of the transition states were generated with Zn incorporated(see Figure 5-2). The PT was modelled with two different substrates: dimethyl (4-nitrophenyl) phosphate(Paraoxon), and diethyl (2-oxochromen-7-yl) phosphate(coumarin)(see Figure 5-3).

[1]This is the term used for the method and it was coined by SK.

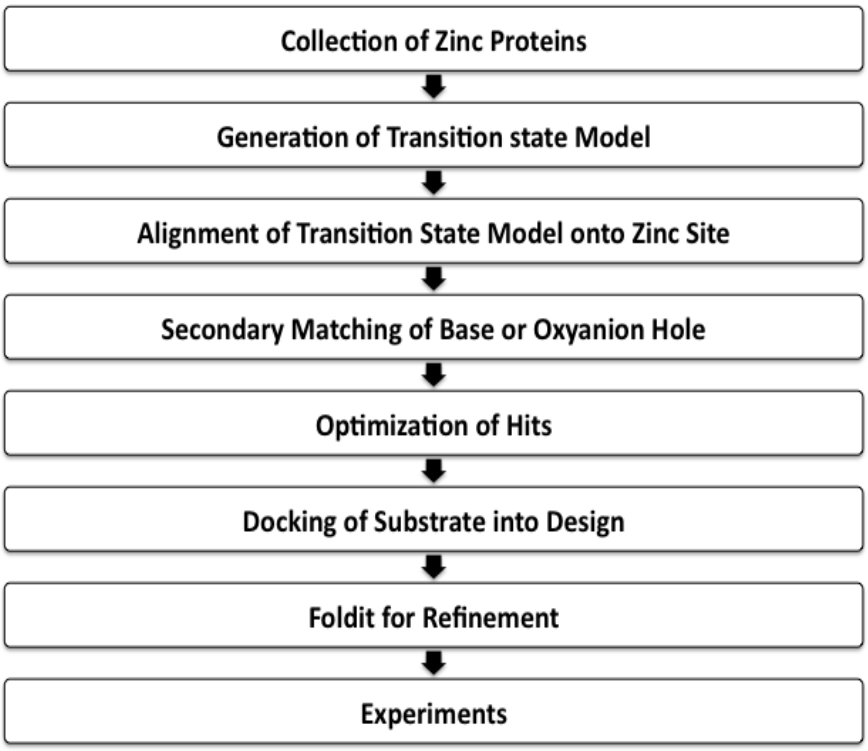
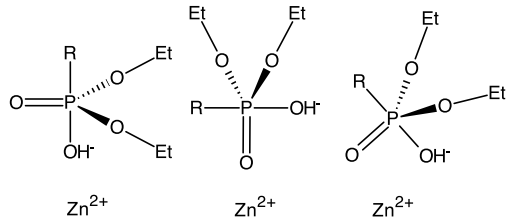


Figure 5-1: Protocol of the Repurposing of Enzymes. Protein scaffolds are collected with Zn. Next, a model of the transition state(TS) is generated. The TS model is aligned to the Zn site in different fashions: monodentate or bidentate. A secondary match is performed for catalytic important residues, a base for the attacking water or stabilization of the oxyanion hole. The hits are optimized for interaction and stabilization using Rosetta. The designed rotamers are evaluated by repacking the designed structure without the substrate to see if they keep their potentials. The TS model is docked into the designed protein to investigate its potential energy surface. The potential candidates are evaluated using the graphical interface Foldit to further optimize the interactions. All potential candidates are then tested experimentally.

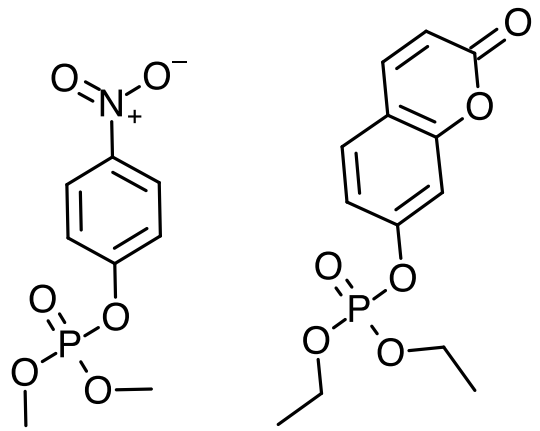


(a) (b) (c)

Figure 5-2: Alignment of Transition State Models. Three different alignments are performed (a) The hydroxyl alignment Zn is used to activate the water molecule. (b) In the phosphoryl alignment Zn is used as a oxyanion hole. (c) In the bidentate alignment Zn is used to lower the pKa of the water and to stabilize the oxyanion hole. Here the model is rotated around the Zn-P axis to search for as many interactions as possible. In all TS models, the leaving group, R, and the hydroxyl are placed in an in-line attack(opposite each other) .

To generate the TS models empirical knowledge was used for angles and distances in the TS. The distance between the coordinating oxygen atom from either the hydroxyl or the phosphoryl oxygen(phosphoryl) and the zinc ion was specified to 2 Å which is close to an average value observed for Zn-water interactions[7]. In the bidentate TS model, the distance is 3 Å between the phosphate atom and Zn. The phosphate atom was hybridized as trigonal bipyramidal, a pentacoordinated fashion, which is characteristic for a nucleophilic attack on the phosphor atom[24]. The attacking hydroxyl was placed in an in-line attack(opposite the leaving group) of the leaving group of the PT, which has been shown to have the lowest energy barrier for the reaction[34].

To increase sampling of the ligand-protein interaction, rotamers of the TS model were generated defining dihedral angles around degrees of freedom that did not affect the geometry of the TS. Intervals and steps sizes were defined and structures were generated combining the different permutations.



(a) (b)

Figure 5-3: The Phosphotriesters Paraoxon and Coumarin. (a) dimethyl (4-nitrophenyl) phosphate(Paraoxon). (b) diethyl (2-oxochromen-7-yl) phosphate(coumarin).

5.2.3 Automated Alignment of Transition State Model onto Native Zinc Site

An algorithm was developed to align the TS model onto the Zn in the native protein scaffold. It used the metal site in the protein and the coordinating atoms to identify the direction of the enzymatic pocket. To explore as many possibilities for the chemical reaction, different alignments were performed: monodentate with the hydroxyl or the phosphoryl coordinating the Zn and a bidentate alignment with both the hydroxyl and phosphoryl aligning the Zn. Depending on the alignments, the heteroatom coordinating the Zn was used to superimpose the TS model.

Constraints were generated from the crystal structures to keep the Zn in the same position during the design process. To avoid major clashed with the protein, the ligand-protein interaction was optimized using the generated rotamer library where all the amino acids of the protein were converted into alanines. Other repulsions were minimized using the Lennard-Jones potential.

5.2.4 Secondary Matching

To rationalize the design process, the secondary matching algorithm implemented in Rosetta was used. Briefly, it goes through all the positions on a protein scaffold to see if it can match the constraints providing by the user.
Here, a secondary match was performed for either a base or an oxyanion hole depending on the suggested reaction mechanism.

5.2.5 Protein Design of Matches

All hits from the secondary matches contained one modification to their sequences either as a base, an oxyanion hole, or positioning of the attacking hydroxyl. To stabilize the modifications introduced, the energy function in Rosetta for protein[103] and ligand[130] was used. The interaction between ligand and protein was optimized for interaction as well as to stabilize the protein sequence for folding. All designs having an interface energy $<$ -3 Rosetta units(RU), solvent accessible surface area(SASA) between 0.8 and $<$ 1 Å^2, and a constraint energy less 3 RU were collected for further analysis.

5.2.6 Evaluation of Mutations

To evaluate the mutations introduced by Rosetta, a multiple sequence alignment was generated to evaluate how conserved the mutated residues were. This was performed to remove substitutions that might be important for folding. All designed sequences were evaluated using the ConSurf server[10, 108]. Residues having a high conservation score and were not part of the native catalytic machinery, were reverted back to their native residue as they were assumed important for folding or solubility. In general, glycine(Gly) and proline(Pro) residues were reverted back to the wild type residues in the designed protein. The designs were evaluated using the visual interface to Rosetta - Foldit[32].
The TS model was docked into the protein to observe if other local energy minima existed and has shown to test if substrates bind[67]. The TS model was docked

into the design structure 10.000 times using the docking algorithm implemented in Rosetta[130]. If other local binding minima existed the binding of the TS model was reoptimized.

5.2.7 Molecular Dynamics Simulation of Designs

To explore the phase space of the substrate and get statistics on different angles involved in the designed structure a MD simulation was performed. The hydrogen bond(HB) network for histidine(His) residues was established by visual inspection of the proteins(native and designed protein). The following His were protonated on Nϵ: 105, 157, 258, 278 and the rest were protonated on Nδ: 15, 17, 138, 197, 210, 214, 238, 241. The parameters for the substrate were generated using the module Antechamber[198] within AmberTools[26]. The metal site was fixed by applying bond, angle, and dihedral constraints to the site. The weights were relatively high and the charge of the metal ion was set to +2. The structures were minimized for 10.000 steps with 5.000 steps using SD and 5.000 with GC. The complexes were solvated and neutralized which gave a total of 12330 water molecules in the complex without substrate and 11414 water molecules in the complex with substrate. The water molecules were parametrized as TIP3P[90]. The whole system was minimized using 5.000 steps of SD and 5.000 of CG. After minimization, the system was heated up to 300 K using a weak restraint on the protein complex, 10 kcal/mol, for 100 ps using an integration step of 1 fs. The volume was fixed for equilibration of the pressure. Next, the system was equilibrate for 1 ns in the NPT ensemble keeping a pressure of 1 atm and the temperature constant at 300 K using a Langevin thermostat with a collision frequency of 1 ps^{-1} and the integration was changed to 2 fs. All hydrogens were fixed using SHAKE[161] and full electrostatics were computed using the Particle Mesh Ewald[35] with periodic boundary conditions. The MD simulation was run for 20 ns using the ff99SB force field[76].

5.3 Results

12 designs were generated based on the protocol described. I will shortly describe two of the designs from which general issues regarding the design process can be illustrated.

5.3.1 Redesign of Adenine Deaminase

One of the designs is a adenine deaminase(ADA)(PDB ID: 1A4L) where 8 mutations have been designed into scaffold. Figure 5-4 shows the design along with the most important residues along with the substrate.
From the secondary matching, a glutamine(Gln) is placed to stabilize the position of the water molecule in the phosphoryl alignment of the TS model. The base for the reaction is believed to be the glutamic acid(Glu) 217 which is a native residue. The rest of the mutations can be divided into two groups

- Making room for the substrate to fit into the pocket
- Packing of the transition state

The following residues Asp19Ser, Phe61Thr, Leu62Ile, Asp296Ala are mutated to make room for the substrate to fit into the pocket of the protein.
Asp19Ser decreases the penalty of desolvation which has to been paid for the substrate to the enter the binding pocket. The mutation makes room for the 4-nitrophenyl group where the Asp19 would have made a electrostatic as well as a van der Waals(vdw) repulsion with the leaving group. Phe61Thr makes room for the methyl ether group while the Leu62Ile makes room for the leaving group. The Asp296Ala mutation is necessary to fit the methyl ether near the Zn. This mutation is important for the interaction as it removes a vdw repulsion as well as a big desolvation penalty.
The Phe65Trp makes a π-π interaction with the leaving group of the substrate. The mutation is supported by a second shell change, Ile299Val, which makes room for the bigger indole ring of Trp. The Ala183Ile is introduced opposite and packs the

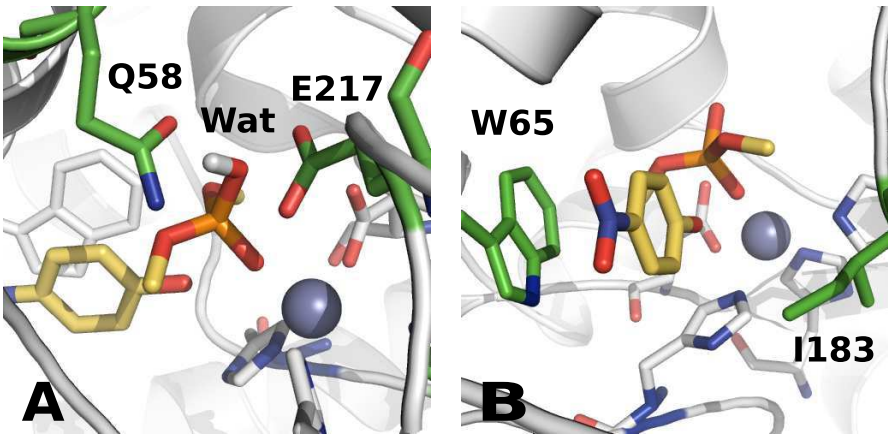


Figure 5-4: Phosphotriesterase Design of Adenine Deaminase. (a) shows the Q58 from the secondary matching which positions the attacking hydroxyl by its carbonyl atom. It is observed that depending on the alignment of the substrate it is possible to stabilize the phosphoryl by a rotation of the head group such that the amine could stabilize a negative charge during the reaction mechanism. (b) shows the mutations introduced for optimization of the TS binding. The W65 makes a π-π interaction with the leaving group.

leaving group by vdw interactions. The designed enzyme has potential to catalyse the reaction since it has a good catalytic machinery along with nice binding interactions.

5.3.2 Redesign of Unknown Function from the Glyoxalase Family

In this design, 11 mutations are inserted into the protein scaffold to transform it into a potential PTE. The alignment of the substrate is performed in a bidentate fashion. From the secondary match, an Asp is matched as a base for the reaction replacing Ser61. The designed rotamer has low Boltzmann weight and is not backup as seen in native enzymes. This is problematic but it is not possible to support it without major modifications of the protein backbone.

Mutations are introduced to bind the substrate as well as prevent clashes between enzyme and substrate. The leaving group is polarised by a Lys residue at position, His46Lys, which is backup by a cation-π interaction by a designed phenylalanine at position 39(Gln39Phe).

Table 5.1: Kinetic Values for Phosphotriester Reaction

PDB ID	$k_{cat}(1/sec)*10^{-4}$	K_M	$k_{cat}(1/sec)/K_M(sec*M)$
1A4L	N.D.[1]	N.D.	N.D
PT3	2	30	7
$PT3_{tm}$[2]	41	29	140
$PT3_{tm}$ E217Q	0.8	43	2
$PT3_{tm}$ Q58H	11	29	38
$PT3_{tm}$ Q58L	73	33	220
$PT3_{tm}$ Q58S, T268R	N.D.	N.D.	N.D.
$PT3_{tm}$ Q58S, T268K	N.D.	N.D.	N.D.

[1] Non Detectable
[2] Is a triple mutant I62L, V218F, V299E

5.3.3 Summary of Experiment

From the designed sequences, one of them shows activity - the repurposing of the adenine deaminase(PDB ID 1A4L). For the activity assay, the coumarin substrate has been used to measure the activity as it is not as labile as the paraoxon. The substrate has been docked into the designed structure and fits well.

A library scan is performed around the active site of the enzymes by trying all 20 amino acids at different positions. From the results, a triple mutant is isolated, Ile62Leu, Val218Phe, Val299Glu where Ile62Leu is a mutation back to the wild type. The Val299Glu is a second shell mutation which stabilizes the helix and especially the Trp65 substitution. The Val218Phe seems to spatially interact with Glu217 and Thr269 and could indirectly stabilize an internal HB network by restricting these two residues. Gln58 does not contribute as suggested by the design where it should have placed the hydroxyl for nucleophilic attack or functioned as an oxyanion for the phosphoryl.

Mutations are performed to see if Glu217 is an oxyanion hole or used to place the hydroxyl. The Glu217Gln show an increase in K_M which is probably due to easier desolvation of the pocket, but k_{cat} decreased significantly.

A crystal structure of the protein is solved and as expected from the kinetic data the Gln58 is solvated and has flipped out of the pocket. A loop(residue 217-220) is displaced due to the binding of a calcium ions but otherwise there is an overall agreement with the designed protein.

Table 5.2: Sequence Information For Phosphotriesterase

PDB ID	Mutations	Function	Superfamily
1A4L	WT	Adenosine Deaminase	Amidohydrolases
PTE(Design)	D19S,L58Q,F61T,L62I, F65W,A183L,D296A,I299V	Phosphotriesterase	Amidohydrolases

5.3.4 Molecular Dynamic Simulation

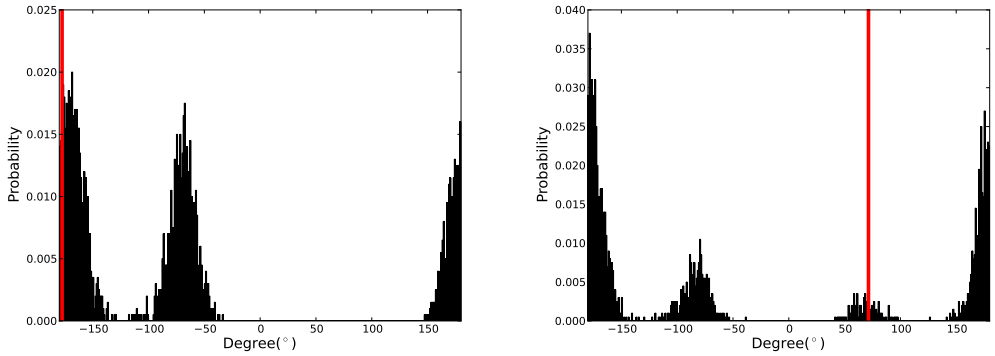
To investigate the equilibrium distribution of the χ_1 and χ_2-angle of Gln58, a MD simulation is performed. The displacement of the loop, 217-220, is investigated to determine if it is crystal artifact induced by the calcium ion.

Figure 5-5 shows the distributions of the χ_1- and χ_2-angle with and without coumarin in the simulation. The χ_1 angle has its density in the trans-configuration(-180° or 180°) both with and without coumarin. The χ_2-angle has its main density in the trans-configuration and in the simulation without substrate, it has a small population in the gauche(+)-configuration as designed. In the simulation with substrate, there is no population in the gauche(+)-configuration. It is still pointing inwards and is support by a HB network of water. During the simulations it is not observed that the Gln58 flips out of the lumen as in the crystal structure. The loop displaced by the calcium ion in the crystal structure goes back to the same position as in the wild type(WT) indicating it is displaced by crystallization artifacts.

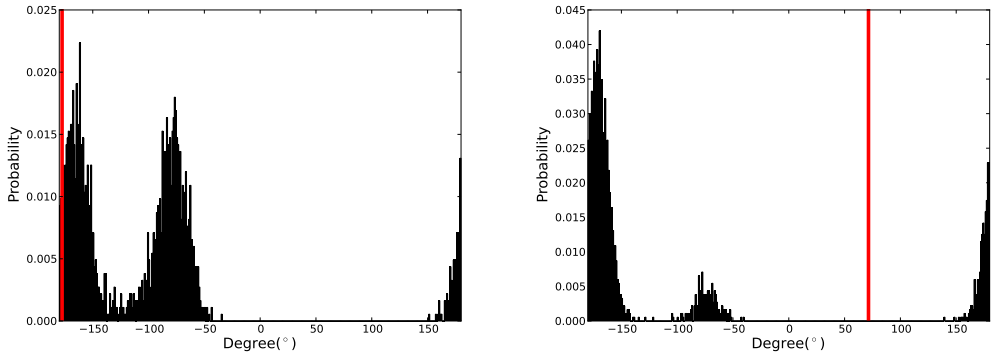
The fluctuations in the active site are monitored and by introducing the substrate into the simulation rather large fluctuations are observed for the supporting helix(residue 57-69). This could indicate that a further stabilization of the helix can improve the reaction rate.

5.4 Discussion

It has been shown how to computationally repurpose an enzyme to have new function. The designed enzyme shows catalytic activity while the wild type enzyme did not show any activity. The computational filters have been able to capture the most important



(a) χ_1 (b) χ_2



(c) χ_1 (d) χ_2

Figure 5-5: χ_1- and χ_2-angle Distributions of Gln58. The red vertical bar represents the designed values. (a)-(b) show the simulation without any substrate. For the χ_1-angle, the most populated is the same as the designed value whereas in case of the χ_2-angle the designed value has a low population. (c)-(d) show the χ-distributions in the simulation with coumarin. The designed χ_2-value is not sampled indicating it is not a favourable rotamer.

features of the designs and to narrow down the amount of designs into a manageable size for visual inspection as well as experimental testing. It has not been unnoticed that the ADA is part of the Amidohydrolase family and this will be used in chapter 6 to see if one can improve the rates.

5.4.1 Alignment of Substrate in new PTE

The crystal structure of the designed protein shows that not all predictions are in agreement with the computationally predicted ones. In the computationally predicted TS model, the water molecule is stabilized by Gln58 and Zn is used as an oxyanion hole. In the native enzyme, the Zn is used to activate a water molecule. An alternative reaction mechanism is possible in contrast to the designed one. Here, the hydroxyl coordinates the Zn while the phosphoryl is stabilized by an oxyanion hole. In the native enzyme, a hydroxyl is coordinating Zn and makes a nucleophilic attack on adenine[57, 175]. His238 is the base which substracts the proton from the coordinating water. The hydroxyl is fixated by Asp295[204] which is coordinating the Zn. The symmetry of the oxyanion hole and the coordinating water have been considered where one can exchange the position of the phosphoryl and the hydroxyl. In this cases, the Gln58 can rotate and function as an oxyanion hole for the reaction. The orientation of Gln58 has been studied by MD simulations and shows that the designed χ_2-angle is not preferred by the enzyme. Experimental data confirm that Gln58 is not stabilizing the transition state. This is shown by reverting it back to the native residue Leu which increases the k_{cat}.

A lid mechanism has been proposed for the native enzyme upon binding of the substrate[188]. It might be this interaction which is decreased by the mutation but future studies will have to be made. The dynamics of the protein is not possible to model within a static frame. Other techniques such as MD simulations or NMR experiments can be used as a supplement in the design process to determine the importance of fluctuations.

The finding indicates that the reaction rate can be increased further by introducing more mutations to stabilize the scaffold. To conclude how the reaction proceeds more

studies need to be done both experimentally and theoretically.

5.4.2 Expression of Designs

The changes of amino acids around the active site can for some fold classes change the folding pathway radically or even misfold the protein. In this study 6 out of 12 proteins have been expressed. There can be many reasons for this but one is that the folding kinetics or the 3D structures has been perturbed too much and they are not able to recover their native folds. The expression levels in non metallodesigns are even lower compared with these values. It could be speculated that metalloenzymes are more prone to mutations due to the stabilization by the binding of metal ions.

5.4.3 Chemical Rate Improvement by 3 Mutations

The 3 mutations, Ile62Leu, Val218Phe, and Val299Asp, are introduced by library scans and stabilize the protein scaffold. Their main effect is on the catalytic rate of the reaction. The reversion of Ile62Leu which is located on the helix, could be involved in binding the substrate and stabilize it during the reaction.
The Val218Phe mutation is interesting as it screens the Glu217 which is a potential base. The mutation makes a steric fixation of Glu217 and lowers its pKa value. The Glu217Gln mutation shows that it is its ability to substract or donate protons and not only polarize. This indicates that the initial alignment of the substrate might be correct.
Val299Asp stabilizes the helix by a HB with Tyr68 where Trp65 has been placed. The three introduced mutations are hard to predict computationally and illustrates one of the main limitations in the computational design. The mutations stabilize the scaffold and are second or third shells from the catalytic residues.

5.4.4 Too Flexible Catalytic Residues

Issues related to the design process have been raised by this study. In some of the designs second and third shell support of the catalytic machinery has not been

captured by the design algorithm. This might explain why some of the designs have not shown any activity. The support of catalytic residues are seen in native enzymes where catalytic residues have low configurational entropy[85]. The second and third shell stabilization is not trivial. It would require remodelling of the protein scaffold changing either secondary structure(SS) elements or loops. In the design, Gln58 is not backup by steric interactions or a stable HB network. For future designs, one should go for more favourable rotamers and try to stabilize the designed residues more. Optimization of catalytic residues can require extensive remodelling of the surrounding loops and SS elements, but can be done within the frame work of Rosetta.

5.4.5 Pitfalls of Computational Repurposing

The design of metalloenzymes is extremely challenging and still much progress has to be made[121]. There can be many reasons for the failures of the designed enzymes and rates are still far below diffusion limit[13] - here are mentioned a few

- wrong modelling of the chemical reaction
- design can not reach its active state
- the protein matrix does not support the mutations

In the case of PTE design the chemical reaction is simple and proceeds through hydrolysis. The Zn provides either an oxyanion hole or lowers the pKa of the water and thereby activates it. The alignment can be problematic as seen from the above but symmetrical properties decreases the error in this reaction. The generation of other parameters for the TS model needs to be considered as well. The main point to be raised is that if the TS model is wrongly aligned or constructed, the computational design will not work.

The energetics of the designed complexes can introduce errors which will give inactive enzymes. The chosen rotamers for the chemical reaction can be high energy conformations which are not possible to reach through thermal fluctuations. This will also

lead to inactive enzymes.

The mutations introduced to stabilize the interaction between protein and ligand, might not be supported by the scaffold. The importance of the stability of the scaffold is shown in this study, as many of the mutations seem to stabilize the scaffold for the interaction with the ligand. It is shown by the Val299Asp mutation how much stabilization of the scaffold increases activity. Computationally it is still beyond our approach to predict these distal changes. The main point is that much care has to be taken into account during the design process as much can go wrong.

5.4.6 Improvements for Repurposing of Metalloenzymes in Rosetta

Electrostatics for Metal ions

Electrostatics is a big issues in protein design[197]. A simple model is used to compute the electrostatics and for nonmetal enzymes it might be reasonable. For metal ions, the approximation becomes too simplistic as charge transfer and polarization effects are large effects for ions. The Generalized Born(GB) solvation model[141] has been used in DNA/protein designs[11] and for metal designs they might be better described using the GB model.

Even though Zn is often used in nature due to its simple behavior, it has been shown from surface catalysis and catalytic nanoparticles that much catalytic activity can be gained from transition metals[137]. These metal ions behave non-classically with respect to spin and coordination, e.g. lowering of symmetry by Jahn-Teller effects as observed in copper sites[180]. In redox processes, it is necessary to use transition metals and for these reactions a better parametrization should be implemented in Rosetta.

Multistate Design

More states could be added to the design process, i.e. the unfolded state, the bound state, the transition state, and the unbound state. This would optimize the structure

for all the stages during the reaction. The multistate design algorithm[6, 207] has been described and is implemented in Rosetta. The problem is that the computational cost increases by including more states but is manageable for the later stages when fewer potential designs are present.

Remodelling of Backbone Structures

More radical changes of the protein structure is another strategy to increase catalytic efficiency. Remodelling of loops could be performed to support catalytic residues or improve binding of the TS model which has shown some success[134]. Implementing these changes, would create an iterative process for the designs where a design is improved by first side chain interaction followed by scaffold modelling such as loops. This raises new demands to the energy function as more conserved residues are removed or remodelled. By expanding the modelling, the folding and expression problem of the enzyme will become larger and more computational filters are needed to evaluate these changes. From alignments of native proteins catalyzing the same reaction, insertions and deletions of the protein sequence are observed. It is beyond the capabilities to model these at the moment.

Restriction of Catalytic Residues

Force field improvement needs to be performed where core residues are easier to model compared with surface residues. The Rosetta force field has been develop for folding which is mainly modelling of core residues that are important. The residues on surfaces or in pockets are generally involved in the catalysis and improvement in their rotamers and desolvation energies would greatly improve the design process so much still needs to be tested.

5.4.7 Evaluation of Enzyme Design in Rosetta

A question raised by Warshel is if the methodology is too simplistic to design the correct catalytic effect[53]. It is correct that many parameters are estimated in a

simplified way such as solvation and electrostatics as discussed above.

The observed activity in the designed enzymes is due to cage effects where charges are constrained and with some probability they stabilize the TS(For more on cage effects see [196]). Simulations were performed and suggestions to TS stabilization were made which should have increased k_{cat} by 3 kcal/mol, but this was not seen. More suggestions have been made but have not been tested experimentally. The discussion on TS stabilization shows the need to include QM into the design process at some stage. The problem is computational time and complexity by introducing QM.

Other computational techniques besides Rosetta have been used to supplement information of the designed enzyme such as MD simulation[96, 176] and QM/MM[5].

To replace the design algorithm in Rosetta with MD simulations is not a practical approach to improve designs. Sequence space is too large for MD simulation to be able to evaluate the different permutations but to evaluate or rank designs, it is a great technique. For metalloenzymes though, MD simulation can be problematic with respect to the parameters for the metal ion. Often metal ions are treated as a point charge with a van der Waals potential, e.g. in simulations of alcohol dehydrogenase[122], others parametrize the metal and its interaction[162]. Here, the MD simulations have been performed using a point charge and restraints to keep the zinc ion at the right point group.

QM/MM is even more demanding computationally compared with MD. It evaluates the energy barriers for active designs by analysing how the transition state is modelled/stabilized. These simulations give information on how the TS is stabilized by the enzyme. It could show which parts of the TS are needed to be stabilized to decrease the activation energy. With the new PTE, it would be interesting to test if one could determine the reaction mechanism using QM/MM and should be tried.

5.4.8 Perspective of Repurposed Metalloenzymes

As mentioned earlier, the use of metals has great advantages due to their catalytic properties and the stability of metal sites[65]. This makes the method a pragmatic solution for generating new enzymes.

If one is able to obtain activity which can easily be measured, the metalloenzyme is superior to e.g. catalytic antibodies. It is easier to introduce new changes into the enzyme either rationally or through random mutations. As shown above, these techniques improve the enzyme by many factors and start to resemble native enzymes. Catalytic antibodies are in rare instances able to reach native rates[190, 117] but are in general much lower. Catalytic antibodies have been develop for PT reactions and have shown only modest acceleration compared with real enzymes[182, 111] and even compared with the ones developed in this study. From a practical point of view, one is not limited by the ability of synthetic chemistry to mimic the transition state and makes repurposing a more promising technique. Furthermore, to supplement the designs with experimental techniques such as library scans or DE improve designs tremendously as shown here.

Compared with earlier computational de novo designs[89, 159] the starting activity has shown to be higher in the first round of designs for the metalloenzymes. The wrong modelling of the Gln58 could have had a greater impact on the reaction rate using a nonmetal system and might have abolished the activity. The advantage with metalloenzymes is that the metal ion does most of the catalysis. In carbonic anhydrase, the pKa value of the zinc coordinating water molecule is lowered from 15.7 to 7.5[164] whereas to lower the pKa by 7 units using amino acids requires a combination of many side chains[64].

5.4.9 Conclusions

It seems to be a promising strategy to repurpose existing metalloenzymes towards new chemical reactions. Still much improvement can be incorporated into the computational design.

Furthermore, the starting rate of the chemical reaction is descent for future work on the scaffold. Using metalloenzymes is a promising technique and should be tested on other reactions. With the above results and how much the efficiency is improved by library scan much computational work still needs to be performed. The reaction rates are still far from diffusion limits which is the final goal to reach hence much

more work needs to be done in the field of enzyme design.

5.5 Acknowledgements

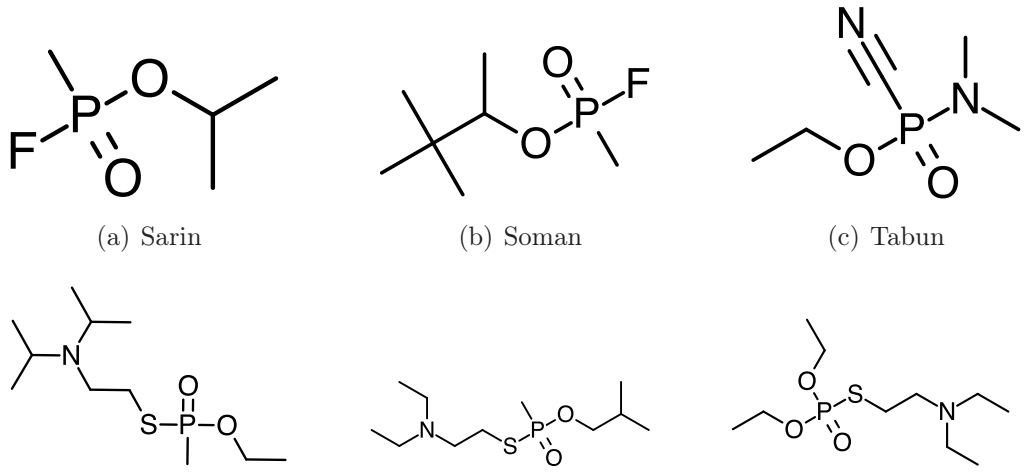
This work was carried out at the Bakerlab at the University of Washington. The research was designed by Per Jr. Greisen(PG), Sagar Khare(SK), and David Baker(DB). All computational work was performed by PG and designs were discussed with SK and DB. The experimental results and later mutagenesis were performed by SK and Yakov Kipnis(YK) in the Bakerlab.

Chapter 6

Computational Redesign of Binuclear Metalloproteins

6.1 Introduction

Organophosphates are toxic compounds that are used as pesticides or chemical warfare agents of which nerve agents are the most deadliests. They bind to an activated serine residue in the enzyme acetylcholinesterase and prevent it from breaking down acetylcholine and thereby paralyze the muscles breathing causing suffocation. Phosphotriesters(PTs) are used in industrial processes and are not easily substituted.
Here, the focus is on nerve agents which are divided into two types, the G- and V-series. The G is an annotation for Germany and the V is for Venomous[27]. Among the G-series are compounds like sarin, soman, and tabun, in the V-series are compounds like VX, russian VX, and VG, also known as amiton(see Figure 6-1)
The compounds can be degraded through alkaline hydrolysis and in case of VX the degradation is performed at 90° C. The reason for the elevated temperature is that at room temperature the chemical reaction will give mixed products some of which are still poisonous[78, 34]. As illustrated above, the decomposition of the compounds is problematic and e.g., VX can be stable for 2-3 months without any notable decomposition. There is a great need to be able to decompose these substances in an easy way. Decomposition of PTs through enzymes would be great solution but few



(a) Sarin (b) Soman (c) Tabun

(d) VX (e) Russian VX (f) Amiton

Figure 6-1: The G- and V-series of Phosphostriesters. Figure (a)-(c) show the G-series of the PTs where sarin and soman contain a fluoride atom while tabun has a triple bonded nitrogen to a phosphorous atom. Figure (d)-(f) show the V-series which have a sulfur-atom instead of a oxygen atom for the leaving group. This increases the pKa-value of the leaving group and decreases the rate of hydrolysis.

enzymes are known to be able to catalyse the P-S bond as in VX. Organophosphorus hydrolase, though, has been able to break the bond[42, 101].

6.2 Phosphotriesterases in Microbacteria

Enzymes have been isolated that have phosphotriesterase(PTE) activity and are able to degrade PTs but their native substrate has not been isolated. It is thought that the development of PTEs has been through evolutionary pressure of microbacteria exposed to pesticides. It has been observed that evolutionary pressures can create new functions[93].

6.2.1 The Superfamily of Amidohydrolases

The isolated enzymes with phosphotriesterase activity all belong to the same protein superfamily, amidohydrolases[158] which consists of a wide range of enzymes: urease, phosphotriesterase, adenosine deaminase, dihydroorotase, and aminoacylase[70]. They all have the characteristic fold, $(\alpha\beta)_8$, which is a common motif among

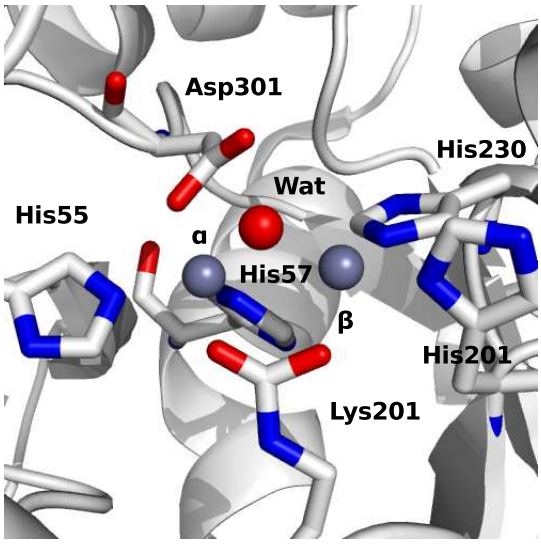


Figure 6-2: The Active Site of Phosphotriesterase from *Pseudomonas Diminuta*. The α and β-sites are marked. The oxygen of the hydroxyl-group is illustrated by a red sphere between the two zinc ions. The carbamylated lysine is coordinating both zinc ions. The α-site is trigonal bipyramidal and the β-site is distorted tetrahedral(PDB ID: 1HZY).

proteins[47]. The catalytic site contains metal ions and depending on the enzyme, it has one or two ions. The type of metal ion can vary depending on the enzymes, but most of them use zinc though other metal ions are found such as nickel in urease. Some of the amidohydrolases have a post-translational modification where a lysine(Lys) residue is carbamylated and is part of the binuclear metal center. The two oxygen from the carbamylated Lys are coordinated to both of the metal ions.

6.2.2 Binuclear Metal Site in Phosphotriesterase

The binuclear Zn sites are divided into an α- and β-site, where the α-site is the more buried(less solvent exposed) while the β-site is solvent exposed. The α-site is coordinated by three residues: His55, His57, Asp301 as well as the oxygen from the hydroxyl and an oxygen from the carbamylated Lys residue. The α-site is trigonal bipyramidal coordinated. The β-site is coordinated to His201, His230, and to the oxygens from hydroxyl and the carbamylated Lys and its geometry is a distorted tetrahedral(see Figure 6-2)[194].

6.2.3 Catalytic Mechanism for Phosphotriesterase

The reaction mechanism for the hydrolysis of a PT starts by binding of the substrate in the enzyme pocket. The phosphoryl oxygen of the substrate binds to the β-site which is the more solvent exposed site. The phosphoryl group is polarised by the zinc ion making the phosphate atom more electrophilic and susceptible for a nucleophilic attack. The nucleophilic attack is performed by the hydroxyl group located between the two metal ions[38]. A base is present to extract the proton from the water molecule(His254 in this example). It has been show that the base is not essential for reactions with low pKa-values of the leaving group e.g., Paraoxon, as the pKa-value of the coordinating water molecule is lowered by the two metal ions. The reaction rate is only lowered by a factor of 2 in the case of Paraoxon while it affects the rate constant more for the leaving groups with higher pKa-value[72]. It has been shown that for leaving groups with pKas > 7, the chemical step is the rate determining step, while < 7, it is the diffusion of the substrate which is the rate limiting stay[38].
During the reaction, the negative charge of the hydroxyl which has been delocalized between the two metal ions, is displaced towards the electrophilic phosphate atom. This creates an electrostatic repulsion between the two metal ions and their distance is increased thereby displacing the backbone[205].
It has been established through theoretical calculations that the metal ions are TBP coordinated during the reaction where the β-site gets an extra oxygen atom from the coordinating substrate[205].The active site is regenerated through a new water molecule that goes in and is deprotonated through a proton shuttle mechanism going from Asp301 -> His254 -> Asp233[12].
It is known that the protein scaffold induces restrictions on the substrate recognition and a fair degree of stereo selectivity is seen for wild type PTEs.

6.3 Aim of the Study

There is a great interest in the use of PTEs for the detoxification of organophosphates such as nerve agents and for the degradation of agricultural, or household pesticides

hence a need for designing new enzymes.

The new designs should be able to hydrolyse both isomers with the same efficiency as the native enzymes seem to have difficulties hydrolysing different enantiomers[73]. For the redesign of the binuclear amidohydrolases, the inspiration is taken from the native PTEs and their mechanism will be designed into the new enzymes. Polarization of the leaving group will be designed into the designs.

6.4 Methods

6.4.1 Collection of Protein Scaffolds

All protein scaffolds in the protein data bank(PDB) which contain metal ions were collected. Protein structures with high resolution(< 2.5 Å) and sequence identity $<$ 90 % were collected. No restrictions were imposed on the expression organism and no restrictions on the metals present in the structure.

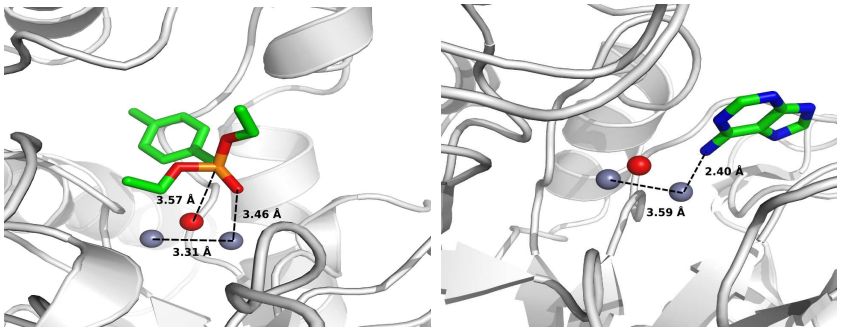
Mononuclear or polynuclear(> 2) metal sites were separated out. A functional binuclear metal site was defined as a metal-metal distance less than 4.0 Å and everything above was discarded.

6.4.2 Generation of Transition State Model

The transition state model was generated for the substrate in the same fashion as for the mononuclear site. For the binuclear zinc sites, knowledge of the reaction mechanism was incorporated into the alignment of the TS model where the metal was differentiated between the α and β site. For other sites, there was no discrimination between the two metal sites.

Rotamers were generated for the TS model around the degrees of freedom not affecting the metal site. The metal site in the TS model was constructed in various ways using different distances between the two metal ions as well as mono and bidentate alignment of the substrate.

An algorithm was developed to extract restraints which were made from the native



(a) (b)

Figure 6-3: Crystal Structures with Bound Ligands. (a) shows the enzymes phosphotriesterase with an analog bound to the active site. The Zn-Zn distance is 3.31 Å which is used in the TS models(PDB ID: 1DPM). (b) In adenine deaminase with adenine bound to the active site, the Zn-Zn distance is larger, 3.59 Å, and a TS model with this distance is generated(PDB ID: 2ICS).

protein structure. The TS models were varied using different metal-metal distances along with geometries from crystal structures where ligands were coordinating the metal site(see Figure 6-3).

An automated alignment algorithm was developed to align the TS models to the metal site in the proteins. For the TS model of amiton, all scaffolds were used while for coumarin only protein scaffolds not showing any phosphotriesterase activity were used.

6.4.3 Design of Binuclear Site

After the alignments, the interaction of the TS models with the proteins were optimized using the energy function of Rosetta for protein[103] and ligand[130].

An energy cut-off was set to filter out structures having greater than -3 Rosetta energy units. The solvent accessible surface area(SASA) for the ligands were used as a filter, where SASA scores less than 0.8 was discarded. Furthermore, any designs with constraint penalties greater than 1 were removed.

For the visual inspection Foldit[32] was used to optimize interactions. The placement of the transition state was evaluated by superimposing it on the heteroatoms present in the crystal structure. In cases, where TS models with longer metal-metal dis-

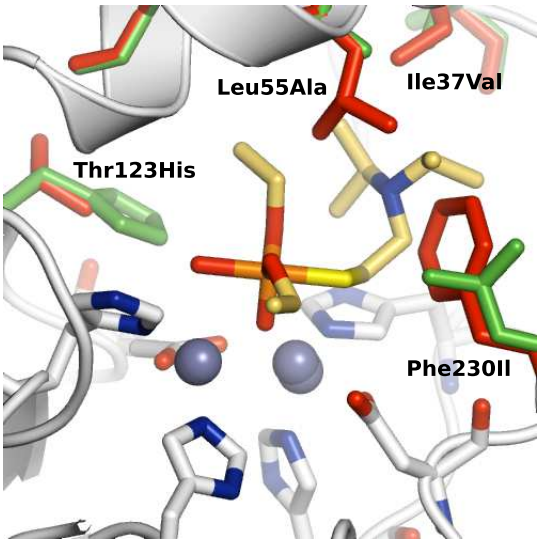


Figure 6-4: Binuclear Enzyme Design with Amiton as Leaving Group. The red colored residues are from the original protein while the green color residues are the substitutions. Most of the substitutions give room for binding the transition state model. An electrostatic interaction is made between the Thr123His nitrogen and the oxygen of the phosphoryl which could stabilize the phosphoryl atom.

tances were used, the designed structure was allowed for to deviate from the imposed superposition with the crystal structure coordinates.

6.5 Results

6.5.1 Amiton

Figure 6-4 shows an example of the amiton design. It has only five mutations, Ile37Val, Leu55Ala, Val58Ile, Thr123His, and Phe230Ile, where four of them mostly make room for the substrate to fit into the binding pocket. The Thr123His is interacting with the substrate. The protein scaffold is an uncharacterized amidohydrolase in the dihydrorotase family. It has to be emphasized that all native scaffolds are tested for activity if any activity is measured in the design structures.

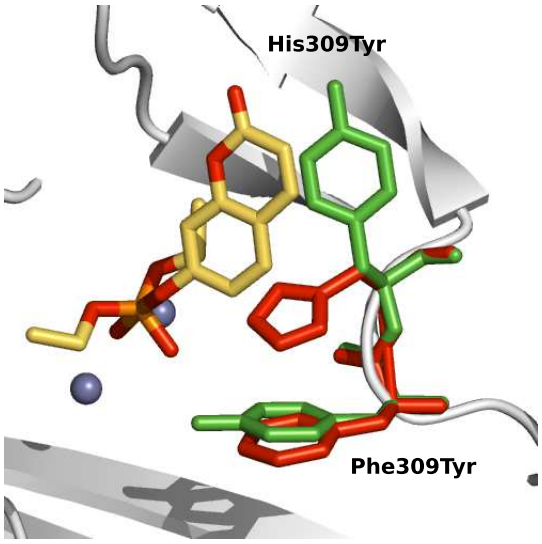


Figure 6-5: Binuclear Enzyme Design with Coumarin as Leaving Group. 11 mutations are introduced into the design and two of them are illustrated. The leaving of group is interacting with the enzymes through a π-π stacking interaction with the His309Tyr mutations. The role of the histidine is not completely clear from the biochemical studies of the native protein but it is not vital for function. To further stabilize the oxyanion hole the F310Y mutation is made.

6.5.2 Coumarin

Figure 6-5 shows one of the designs for coumarin which is a human amino peptidase and the native enzyme contains two cobalt atoms. The protein is an alpha-beta protein and is within the family of alpha-beta complex. 4 mutations has been introduced into the protein scaffold during the design - Cys203Met, Cys301Leu, Phe309Tyr, and His310Tyr. Here, the packing is changed and another kind of stabilization of the TS is tried with the F310Y to stabilize the charge.

6.6 Discussion

The construction of new PTs has been performed with some success using experimental techniques such as directed evolution e.g. the hydrolysis of soman[68]. One has not tried to use computational techniques for optimization of binuclear proteins to hydrolyze PTs.

The sequence space for the binuclear metalloproteins is more limited compared with

the mononuclear metalloproteins described earlier. To bury one charge gives a large entropic penalty and two will have an even higher penalty. This makes the binuclear site much more solvent accessible and less interactions are made from the protein to the substrate[37]. This limits the design process as fewer interactions can be made between the protein and the substrate.

During the design process, loops have been allowed to get displaced in the design of the interaction while secondary structures were not allowed to move during the process. Empirically, it has been shown that by letting loops or secondary structure elements move too much during the design process, the sequences are less prone to express. However, the redesign of the loop, either being an extension or a shortening, will improve the interaction between the protein and the substrate.

To extend the scaffolds to include different metal ions and more than one metal ion has increased the complexity of the design process. More possibilities in the alignment of the TS model have to be tried as more sites are available compared with mononuclear sites.

The lowering of the activation energy by binuclear metalloenzymes is mainly done by the metal ions which activate and stabilise the charges developed during the reaction. Experiments have shown that in some cases it is possible to improve the reaction rate by changing the metal ions present. By using cobalt instead of zinc the specific activity of the PTE was increased[140].

For some substrates, it has been shown experimentally that polarisation or protonation of the leaving group is the rate limiting step for the reaction[38]. It has not been possible to introduce polarization of the leaving group in all designs. In many cases, it has not been possible to place such an interaction into the protein scaffold. This could be improved by a postdesign process where more conserved structural elements are changed such as loops or secondary structural elements.

In the case of the binuclear site there is an electrostatic repulsion between the two positively charged Zn atoms when the hydroxyl group attacks the phosphate atoms. This is shown by theoretical studies where the metal-metal distance approaches 5 Å during the reaction[205].

This distance is argued to be a high estimate and experimental results suggest a metal-metal distance of only 4 Å[95]. Anyway, the main point here is that there is a repulsion during the reaction mechanism between the two metal ions. This has been modelled by varying the metal distances of the TS models.
Questions have been raised concerning how well the designed enzymes do stabilize the transition state and not only make cage effects[53] Much of the this is already done by the metal ions and many of the mutations are mainly for the binding of the TS model.

6.6.1 Outlook and Conclusions

It has been shown that it is possible to computationally change binuclear protein scaffolds and get potential candidates which are similar to native enzymes. Experiments will further confirm if the structures are relatively stable towards mutation and if the all the necessary machinery to catalyze the reactions have been captured by the design process. The final test is of course experiments to see if the new enzymes with binuclear metal sites are generated and whether the actually catalyze as predicted.

6.7 Acknowledgements

This research was design by PG, SK, and DB. The research was performed by PG and results were evaluated with SK and DB. No experiments are yet performed.

Chapter 7

Conclusions

Computational simulations have often lead to a deeper understanding and have assisted in the analysis of experimental work. The algorithms are constantly improved along with force field corrections to obtain better accuracy[120]. Today, it is possible to fold proteins by simulating full all atom proteins on the millisecond time scale[171] which makes the simulations much more biologically relevant.

The new kid on the block within computational biology is the design of new proteins with function and that are stable not only under biological conditions but also under industrially important conditions. It is possible to generate new proteins with a completely new sequence[103] and to generate enzymes designed for catalyzing new reactions[159, 176] and even to generate protein-protein interactions[51]. However, as outlined throughout this thesis, there are clearly many challenges left to be solved for all of the goals to be achieved and the future seems bright and promising.

7.1 Structure Activity Relation between Isomers

In chapter 3, a qualitative difference has been shown between two diastereomers using MD simulations. The simulation times were too short to get a quantitative relationship between the structures but with relatively short extensions this can be obtained.

The combination of MD simulations with vibrational spectroscopy is interesting, but

to obtain the vibrations and their intensities are still problematic. It is hard to isolate few structures from a peptide ensemble and expect that they necessarily represent the average properties observed of the experiments. Certainly, the main problem for combining the two methods is the computational sampling and mainly the quantum mechanical calculations of the Hessian. It is only possible now to calculate for a limited number of atoms and structures which limits the practical application.

7.2 Electron Transfer in Nitrite Reductase

In chapter 4, it has been shown that two pathways are possible for ET in Nir. It has been shown that the metal-metal distance is influenced by the redox states of the enzymes. QM calculations showed that the driving force for the electron transfer is a property of the bridge atoms.

The MD simulations have been performed with water coordinating the T2 site and it would be interesting to monitor the changes by replacing it with nitrite. This would fulfill the catalytic cycle from electron transfer to catalysis.

In general, simulations should always be compared with experiments but to access the electronic pathway through experiments is very difficult. The apparent experimental rate constant for the reaction contains the contributions from both pathways. Here, one can deduce could be applied with the data present here.

7.3 The Creation of a Phosphotriesterase

In chapter 5, the method for repurposing of metalloenzymes using Rosetta has been shown. The experimental results have shown the limitations of the methods. For example, only one of the designs worked out of the 12 proposed. This could be due to both expression problems as well as the high energy rotamers which resulted. Certainly, more questions have been raised by the study that need to be answered.

The sampling of low Boltzman weight rotamers in Rosetta has been addressed where much more flexibility is observed for designed residues compared with native

residues[52]. Rigidifying weights have recently been implemented into the Rosetta algorithm which will hopefully lead to more reliable designs.
A reaction mechanism of the new phosphotriesterase has not been established and it could be interesting to use QM/MM simulations to calculate the proton affinities for the catalytic residues as well as the energy barrier.

7.4 Redesign of Binuclear Metalloproteins

In chapter 6, repurposing of binuclear metalloproteins has been performed and the computational design has given promising properties for catalysing this reaction. No experiments have been performed to date hence it is not known how well the designs perform.
In many cases, the protein scaffolds have limited space for modulations since a certain amount of solvent has to be accessible for the metal ions. The space is even more limited compared to the mononuclear metal sites. Here, an iterative process for the designs should be performed and switching between sequence space to loop remodelling. This could increase the number of designs that would be potential candidates to catalyse the reaction.

7.5 Summary

In summary, computational simulations have assisted in interpretation of experimental results. In this thesis, I have discussed the prediction of binding structures of peptides, electronic pathways in redox proteins, and lastly, the redesign enzymes to have a new functions.
The increase in high-performance computing has made this area attractive and many interesting results and applications are starting to come from this field in science.

Bibliography

[1] C Adessi and C Soto. Converting a peptide into a drug: strategies to improve stability and bioavailability. *Current medicinal chemistry*, 9(9):963–978, 2002.

[2] E.T Adman, JW Godden, and S Turley. The structure of copper-nitrite reductase from achromobacter cycloclastes at five ph values, with no-2 bound and with type ii copper depleted. *Journal of Biological Chemistry*, 270(46):27458, 1995.

[3] R Ahlrichs, M Bar, M Haser, H Horn, and C Kolmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters*, 162(3):165–169, 1989.

[4] Y.H Ahn. Sustainable nitrogen elimination biotechnologies: A review. *Process Biochemistry*, 41(8):1709–1721, 2006.

[5] A.N Alexandrova, D Röthlisberger, D Baker, and W.L Jorgensen. Catalytic mechanism and performance of computationally designed enzymes for kemp elimination. *Journal of the American Chemical Society*, 130(47):15907–15915, 2008.

[6] BD Allen and SL Mayo. An efficient algorithm for multistate protein design based on faster. *J. Comput. Chem.*, 31(5):904–916, 2010.

[7] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell,

J. R. Rodgers, and D. G. Watson. The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallographica Section B*, 35(10):2331–2339, Oct 1979.

[8] C Andreini, I Bertini, G Cavallaro, G.L Holliday, and J.M Thornton. Metal ions in biological catalysis: from enzyme databases to general principles. *Journal of Biological Inorganic Chemistry*, 13(8):1205–1218, 2008.

[9] F.H Arnold. Combinatorial and computational challenges for biocatalyst design. *Nature*, 409(6817):253–257, 2001.

[10] H Ashkenazy, E Erez, E Martz, T Pupko, and N Ben-Tal. Consurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, 38(Web Server):W529–W533, Jul 2010.

[11] Justin Ashworth, James J Havranek, Carlos M Duarte, Django Sussman, Raymond J Monnat, Barry L Stoddard, and David Baker. Computational redesign of endonuclease dna binding and cleavage specificity. *Nature*, 441(7093):656–659, Jun 2006.

[12] S.D Aubert, Y Li, and F.M Raushel. Mechanism for the hydrolysis of organophosphates by the bacterial phosphotriesterase. *Biochemistry*, 43(19):5707–5715, 2004.

[13] David Baker. An exciting but challenging road ahead for computational enzyme design. *Protein Science*, 19(10):1817–1819, Sep 2010.

[14] A Bar-Even, E Noor, Y Savir, and W Liebermeister. . . . The moderately efficient enzyme: evolutionary and physico-chemical trends shaping enzyme parameters. *Biochemistry*, Jan 2011.

[15] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098–3100, Sep 1988.

[16] A.D Becke. Density-functional thermochemistry. iii. the role of exact exchange. *Chem. Phys*, 98(1):5648–5652, 1993.

[17] DN Beratan, JN Betts, and JN Onuchic. Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science*, 252(5010):1285, 1991.

[18] D.N Beratan, J.N Onuchic, and JJ Hopfield. Electron tunneling through covalent and noncovalent pathways in proteins. *J. Chem. Phys.*, 86:4488, 1987.

[19] M.J Berridge. Neuronal calcium signaling review. *Neuron*, 21:13–26, 1998.

[20] J.D Bloom, M.M Meyer, P Meinhold, C.R Otey, D MacMillan, and F.H Arnold. Evolving strategies for enzyme engineering. *Current opinion in structural biology*, 15(4):447–452, 2005.

[21] M Broccardo, V Erspamer, G.F Erspamer, G Improta, G Linari, P Melchiorri, and PC Montecucchi. Pharmacological data on dermorphins, a new class of potent opioid peptides from amphibian skin. *British Journal of Pharmacology*, 73(3):625, 1981.

[22] B.R Brooks, R.E Bruccoleri, and B.D Olafson. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[23] Eric M Brustad and Frances H Arnold. Optimizing non-natural protein function with directed evolution. *Current Opinion in Chemical Biology*, 15(2):201–210, Apr 2011.

[24] S.R Caldwell, F.M Raushel, P.M Weiss, and WW Cleland. Transition-state structures for enzymic and alkaline phosphotriester hydrolysis. *Biochemistry*, 30(30):7444–7450, 1991.

[25] M Cascella, A Magistrato, I Tavernelli, P Carloni, and U Rothlisberger. Role of protein frame and solvent for the redox properties of azurin from pseudomonas

aeruginosa. *Proceedings of the National Academy of Sciences*, 103(52):19641, 2006.

[26] D.A Case, T.E CHEATHAM III, T Darden, H Gohlke, R Luo, K.M Merz Jr, A Onufriev, C Simmerling, B Wang, and R.J Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668, 2005.

[27] S Chauhan, R D'Cruz, S Faruqi, KK Singh, S Varma, M Singh, and V Karthik. Chemical warfare agents. *Environmental Toxicology and Pharmacology*, 26(2):113–122, 2008.

[28] I. Chorkendorff and J.W Niemantsverdriet. *Concepts of Modern Catalysis and Kinetics*. Wiley-VCH, 2003.

[29] H.E.M Christensen, L.S Conrad, K.V Mikkelsen, M.K Nielsen, and J Ulstrup. Direct and superexchange electron tunneling at the adjacent and remote sites of higher plant plastocyanins. *Inorg. Chem.*, 29(15):2808–2816, 1990.

[30] D.W Christianson and J.D Cox. Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. *Annual review of biochemistry*, 68(1):33–57, 1999.

[31] J.L Cole, D.P Ballou, and E.I Solomon. Spectroscopic characterization of the peroxide intermediate in the reduction of dioxygen catalyzed by the multicopper oxidases. *Journal of the American Chemical Society*, 113(22):8544–8546, 1991.

[32] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, May 2010.

[33] O Crescenzi, P Amodeo, G Cavicchioni, R Guerrini, D Picone, S Salvadori, T Tancredi, and P.A Temussi. δ-selective opioid peptides containing a single aromatic residue in the message domain: An nmr conformational analysis. *J. Pept. Sci.*, 2(5):290–308, 1996.

[34] K.A Daniel, L.A Kopff, and E.V Patterson. Computational studies on the solvolysis of the chemical warfare agent vx. *Journal of Physical Organic Chemistry*, 21(4):321–328, 2008.

[35] T Darden, D York, and L Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.

[36] J.R Desjarlais and N.D Clarke. Computer search algorithms in protein modification and design. *Current opinion in structural biology*, 8(4):471–475, 1998.

[37] K.A Dill, S.B Ozkan, M.S Shell, and T.R Weikl. The protein folding problem. *Annual review of biophysics*, 37:289, 2008.

[38] W.J Donarski, D.P Dumas, D.P Heitmeyer, V.E Lewis, and F.M Raushel. Structure-activity relationships in the hydrolysis of substrates by the phosphotriesterase from pseudomonas diminuta. *Biochemistry*, 28(11):4650–4655, 1989.

[39] R.O Dror, M.Ø Jensen, D.W Borhani, and D.E Shaw. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *The Journal of General Physiology*, 135(6):555, 2010.

[40] B.J Druker and N.B Lydon. Lessons learned from the development of an ab1 tyrosine kinase inhibitor for chronic myelogenous leukemia. *Journal of Clinical Investigation*, 105(1):3–8, 2000.

[41] B.J Druker, S Tamura, E Buchdunger, S Ohno, G.M Segal, S Fanning, J Zimmermann, and N.B Lydon. Effects of a selective inhibitor of the abl tyrosine kinase on the growth of bcr-abl positive cells. *Nature Medicine*, 2(5):561–566, 1996.

[42] D.P Dumas, H.D Durst, W.G Landis, F.M Raushel, and J.R Wild. Inactivation of organophosphorus nerve agents by the phosphotriesterase from pseudomonas diminuta. *Archives of biochemistry and biophysics*, 277(1):155–159, 1990.

[43] R.L Dunbrack. Backbone-dependent rotamer library for proteins application to side-chain prediction. *Journal of molecular biology*, 230(2):543–574, 1993.

[44] C.L Dupont, S Yang, B Palenik, and P.E Bourne. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proceedings of the National Academy of Sciences*, 103(47):17822, 2006.

[45] S Duquesne, D Destoumieux-Garzón, J Peduzzi, and S Rebuffat. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat. Prod. Rep.*, 24(4):708–734, 2007.

[46] G Falconieri Erspamer and C Severini. Guinea-pig ileum (gpi) and mouse vas deferens (mvd) preparations in the discovery, discrimination and parallel bioassay of opioid peptides. *Pharmacological Research*, 26(2):109–121, 1992.

[47] G.K Farber. An [alpha]/[beta]-barrel full of evolutionary trouble. *Current opinion in structural biology*, 3(3):409–412, 1993.

[48] O Farver, R.R Eady, Z.H.L Abraham, and I Pecht. The intramolecular electron transfer between copper sites of nitrite reductase: a comparison with ascorbate oxidase. *FEBS letters*, 436(2):239–242, 1998.

[49] O Farver, R.R Eady, G Sawers, M Prudêncio, and I Pecht. Met144ala mutation of the copper-containing nitrite reductase from alcaligenes xylosoxidans reverses the intramolecular electron transfer. *FEBS letters*, 561(1-3):173–176, 2004.

[50] K Filip, M Oleszczuk, D Pawlak, J Wójcik, N.N Chung, P.W Schiller, and J Izdebski. Potent side-chain to side-chain cyclized dermorphin analogues containing a carbonyl bridge. *J. Pept. Sci.*, 9(10):649–657, 2003.

[51] S. J Fleishman, T. A Whitehead, D. C Ekiert, C Dreyfus, J. E Corn, E.-M Strauch, I. A Wilson, and D Baker. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, 332(6031):816–821, May 2011.

[52] Sarel J Fleishman, Sagar D Khare, Nobuyasu Koga, and David Baker. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Science*, 20(4):753–757, Mar 2011.

[53] M.P Frushicheva, J Cao, Z.T Chu, and A Warshel. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proceedings of the National Academy of Sciences*, 107(39):16869, 2010.

[54] H Fukunishi, O Watanabe, and S Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.*, 116:9058, 2002.

[55] J.N Gehlen, I Daizadeh, AA Stuchebrukhov, and RA Marcus. Tunneling matrix element in ru-modified blue copper proteins: pruning the protein in search of electron transfer pathways. *Inorganica Chimica Acta*, 243(1-2):271–282, 1996.

[56] S Ghosh, X Xie, A Dey, Y Sun, C.P Scholes, and E.I Solomon. Thermodynamic equilibrium between blue and green copper sites and the role of the protein in controlling function. *Proceedings of the National Academy of Sciences*, 106(13):4969, 2009.

[57] MP Gleeson, NA Burton, and IH Hillier. The mechanism of adenosine deaminase catalysis studied by qm/mm calculations: The role of histidine 238 and the activity of the alanine 238 mutant. *Physical Chemistry Chemical Physics*, 5(19):4272–4278, 2003.

[58] S Gnanakaran and AE García. Helix-coil transition of alanine peptides in water: Force field dependence on the folded and unfolded structures. *Proteins: Structure, Function, and Bioinformatics*, 59(4):773–782, 2005.

[59] JW Godden, S Turley, D.C Teller, E.T Adman, MY Liu, WJ Payne, and J LeGall. The 2.3 angstrom x-ray structure of nitrite reductase from achromobacter cycloclastes. *Science*, 253(5018):438, 1991.

[60] H.B Gray. Long-range electron-transfer in blue copper proteins. *Chemical Society Reviews*, 15(1):17–30, 1986.

[61] H.B Gray and J.R Winkler. Electron transfer in proteins. *Annual review of biochemistry*, 65:537–561, 1996.

[62] JM Guss, HD Bartunik, and HC Freeman. Accuracy and precision in protein structure analysis: restrained least-squares refinement of the structure of poplar plastocyanin at 1.33 a resolution. *Acta Crystallographica Section B: Structural Science*, 48(6):790–811, 1992.

[63] R.L Harris, R.R Eady, S Samar Hasnain, and R Gary Sawers. Coordinate synthesis of azurin i and copper nitrite reductase in alcaligenes xylosoxidans during denitrification. *Arch Microbiol*, 186(3):241–249, 2006.

[64] T.K Harris and G.J Turner. Structural basis of perturbed pka values of catalytic groups in enzyme active sites. *Iubmb Life*, 53(2):85–98, 2002.

[65] H.W Hellinga. Metalloprotein design. *Current opinion in biotechnology*, 7(4):437–441, 1996.

[66] H.W Hellinga and F.M Richards. Construction of new ligand binding sites in proteins of known structure* 1:: I. computer-aided modeling of sites with pre-defined geometry. *Journal of molecular biology*, 222(3):763–785, 1991.

[67] J.C Hermann, E Ghanem, Y Li, F.M Raushel, J.J Irwin, and B.K Shoichet. Predicting substrates by docking high-energy intermediates to enzyme structures. *Journal of the American Chemical Society*, 128(49):15882–15891, 2006.

[68] C.M Hill, W.S Li, J.B Thoden, H.M Holden, and F.M Raushel. Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site. *Journal of the American Chemical Society*, 125(30):8990–8991, 2003.

[69] P Hohenberg and W Kohn. Inhomogeneous electron gas. *Phys. Rev*, 136(3B):B864–B871, 1964.

[70] L Holm and C Sander. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins Structure Function and Genetics*, 28(1):72–82, 1997.

[71] RH Holm, P Kennepohl, and EI Solomon. Structural and functional aspects of metal sites in biology. *Chem. Rev*, 96(7):2239–2314, 1996.

[72] S.B Hong and F.M Raushel. Metal-substrate interactions facilitate the catalytic activity of the bacterial phosphotriesterase. *Biochemistry*, 35(33):10904–10912, 1996.

[73] S.B Hong and F.M Raushel. Stereochemical constraints on the substrate specificity of phosphotriesterase. *Biochemistry*, 38(4):1159–1165, 1999.

[74] JJ Hopfield. Electron transfer between biological molecules by thermally activated tunneling. *Proceedings of the National Academy of Sciences of the United States of America*, 71(9):3640, 1974.

[75] S Horikawa, T Takai, M Toyosato, H Takahashi, M Noda, H Kakidani, T Kubo, T Hirose, S Inayama, and H Hayashida. Isolation and structural organization of the human preproenkephalin b gene. 1983.

[76] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, Nov 2006.

[77] M.J Humphrey and P.S Ringrose. Peptides and related drugs: a review of their absorption, metabolism, and excretion. *Drug metabolism reviews*, 17(3-4):283–310, 1986.

[78] RL Irvine, SS Haraburda, and C Galbis-Reig. Combining sbr systems for chemical and biological treatment: the destruction of the nerve agent vx. *Water science and technology: a journal of the International Association on Water Pollution Research*, 50(10):11, 2004.

[79] H Irving.... 637. the stability of transition-metal complexes. *J. Chem. Soc.*, Jan 1953.

[80] H IWASAKI and T MATSUBARA. A nitrite reductase from achromobacter cycloclastes. *Journal of Biochemistry*, 71(4):645, 1972.

[81] Christian Jãckel and Donald Hilvert. Biocatalysts by evolution. *Current opinion in biotechnology*, 21(6):753–759, Dec 2010.

[82] F Jacobson, A Pistorius, D Farkas, W De Grip, O Hansson, L Sjolin, and R Neutze. ph dependence of copper geometry, reduction potential, and nitrite affinity in nitrite reductase. *Journal of Biological Chemistry*, 282(9):6347–6355, Dec 2006.

[83] W. P. Jencks. Binding-energy, specificity, and enzymic catalysis - circe effect. *ADVANCES IN ENZYMOLOGY AND RELATED AREAS OF MOLECULAR BIOLOGY*, 1975.

[84] WP Jencks. Mechanism of enzyme action. *Annual review of biochemistry*, 32(1):639–676, 1963.

[85] W.P Jencks. On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences of the United States of America*, 78(7):4046, 1981.

[86] K.P Jensen. Computational studies of modified [fe3s4] clusters: Why iron is optimal. *Journal of inorganic biochemistry*, 102(1):87–100, 2008.

[87] KP Jensen and M Rykær. The building blocks of metallothioneins: heterometallic zn2+ and cd2+ clusters from first-principles calculations. *Dalton Trans.*, 2010.

[88] H.Å Jenssen, C.D Fjell, A Cherkasov, and R.E.W Hancock. Qsar modeling and computer aided design of antimicrobial peptides. *J. Pept. Sci.*, 14(1):110–114, 2008.

[89] L Jiang, E. A Althoff, F. R Clemente, L Doyle, D Rothlisberger, A Zanghellini, J. L Gallaher, J. L Betker, F Tanaka, C. F Barbas, D Hilvert, K. N Houk, B. L Stoddard, and D Baker. De novo computational design of retro-aldol enzymes. *Science*, 319(5868):1387–1391, Mar 2008.

[90] W.L Jorgensen, J Chandrasekhar, J.D Madura, R.W Impey, and M.L Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926, 1983.

[91] W Kabsch and C Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[92] JOA De Kerpel and U Ryde. Protein strain in blue copper proteins studied by free energy perturbations. *Proteins: Structure, Function, and Bioinformatics*, 36(2):157–174, 1999.

[93] O Khersonsky, C Roodveldt, and D.S Tawfik. Enzyme promiscuity: evolutionary and mechanistic aspects. *Current Opinion in Chemical Biology*, 10(5):498–508, 2006.

[94] Olga Khersonsky, Daniela Röthlisberger, Orly Dym, Shira Albeck, Colin J Jackson, David Baker, and Dan S Tawfik. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series. *Journal of molecular biology*, 396(4):1025–1042, Mar 2010.

[95] J Kim, P.C Tsai, S.L Chen, F Himo, S.C Almo, and F.M Raushel. Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterase. *Biochemistry*, 47(36):9497–9504, 2008.

[96] G Kiss, D Röthlisberger, D Baker, and KN Houk. Evaluation and ranking of enzyme designs. *Protein Science*.

[97] A Klamt and G Schuurmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, (5):799, Jan 1993.

[98] Wolfram Koch and Max C. Holthausen. *A Chemist's Guide to Density Functional Theory*. Wiley-VCH, Weinheim, 2001.

[99] R.L Koder, J.L.R Anderson, L.A Solomon, K.S Reddy, C.C Moser, and P.L Dutton. Design and engineering of an o2 transport protein. *Nature*, 458(7236):305–309, 2009.

[100] W Kohn and LJ Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev*, 140(4A):A1133–A1138, 1965.

[101] J.E Kolakowski, J.J Defrank, SP Harvey, L.L Szafraniec, W.T Beaudry, K Lai, and J.R Wild. Enzymatic hydrolysis of the chemical warfare agent vx and its neurotoxic analogues by organophosphorus hydrolase. *Biocatalysis and Biotransformation*, 15(4):297–312, 1997.

[102] HW Kosterlitz, JAH Lord, SJ Paterson, and A.A Waterfield. Effects of changes in the structure of enkephalins and of narcotic analgesic drugs on their interactions with $\hat{\text{I}}\frac{1}{4}$ and $\hat{\text{I}}$´-receptors. *British Journal of Pharmacology*, 68(2):333, 1980.

[103] B Kuhlman. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649):1364–1368, Nov 2003.

[104] M Kukimoto, M Nishiyama, M Tanokura, E.T Adman, and S Horinouchi. Studies on protein-protein interaction between copper-containing nitrite reductase and pseudoazurin from alcaligenes faecalis s-6. *Journal of Biological Chemistry*, 271(23):13680, 1996.

[105] P Kupser, K Pagel, J Oomens, N Polfer, B Koksch, G Meijer, and G Helden. Amide-i and-ii vibrations of the cyclic [beta]-sheet model peptide gramicidin s in the gas phase. *Journal of the American Chemical Society*, 132(6):2085–2093, 2010.

[106] Alexander M Kuznetsov and Ulstrup Jens. *Electron Transfer in Chemistry and Biology: An Introduction to the Theory*. Wiley-VCH, 1999.

[107] L.B LaCroix, S.E Shadle, Y Wang, B.A Averill, B Hedman, K.O Hodgson, and E.I Solomon. Electronic structure of the perturbed blue copper site in nitrite reductase: spectroscopic properties, bonding, and implications for the entatic/rack state. *Journal of the American Chemical Society*, 118(33):7755–7768, 1996.

[108] M Landau, I Mayrose, Y Rosenberg, F Glaser, E Martz, T Pupko, and N Ben-Tal. Consurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research*, 33(Web Server):W299–W302, Jul 2005.

[109] Oliver F Lange, Helmut Grubmüller, and Bert L De Groot. Molecular dynamics simulations of protein g challenge nmr-derived correlated backbone motions. *Angewandte Chemie International Edition*, 44(22):3394–3399, May 2005.

[110] R Langen, J.L Colón, D.R Casimiro, T.B Karpishin, J.R Winkler, and H.B Gray. Electron tunneling in proteins: role of the intervening medium. *Journal of Biological Inorganic Chemistry*, 1(3):221–225, 1996.

[111] B.J Lavey and K.D Janda. Catalytic antibody mediated hydrolysis of paraoxon. *The Journal of Organic Chemistry*, 61(21):7633–7636, 1996.

[112] T Lazaridis and M Karplus. Effective energy function for proteins in solution. *Proteins: Structure, Function, and Bioinformatics*, 35(2):133–152, 1999.

[113] L.H Lazarus, WE Wilson, R De Castiglione, and A Guglietta. Dermorphin gene sequence peptide with high affinity and selectivity for delta-opioid receptors. *Journal of Biological Chemistry*, 264(6):3047, 1989.

[114] C Lee, W Yang, and R.G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785, 1988.

[115] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785–789, Jan 1988.

[116] H Lei, C Wu, Z Wang, and Y Duan. Molecular dynamics simulations and free energy analyses on the dimer formation of an amyloidogenic heptapeptide from human [beta] 2-microglobulin: implication for the protofibril structure. *Journal of molecular biology*, 356(4):1049–1063, 2006.

[117] R.A Lerner, S.J Benkovic, and P.G Schultz. At the crossroads of chemistry and immunology: catalytic antibodies. *Science*, 252(5006):659, 1991.

[118] Cyrus Levinthal. Are there pathways for protein folding? *Journal of Medical Physics*, 65(1):44–45, 1968.

[119] G. Liang, G. Chen, W. Niu, and Z. Li. Factor analysis scales of generalized amino acid information as applied in predicting interactions between the human amphiphysin-1 SH3 domains and their peptide ligands. *Chem Biol Drug Des*, 71:345–351, Apr 2008.

[120] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, pages NA–NA, Jan 2010.

[121] Y Lu, N Yeung, N Sieracki, and N.M Marshall. Design of functional metalloproteins. *Nature*, 460(7257):855–862, 2009.

[122] J Luo and T.C Bruice. Ten-nanosecond molecular dynamics simulation of the motions of the horse liver alcohol dehydrogenase phch2o-complex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):16597, 2002.

[123] AP Lyubartsev, AA Martsinovski, SV Shevkunov, and PN Vorontsov-Velyaminov. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, 96:1776, 1992.

[124] J.K Ma, S Lee, M Choi, G.R Bishop, J.P Hosler, and V.L Davidson. The axial ligand and extent of protein folding determine whether zn or cu binds to amicyanin. *Journal of inorganic biochemistry*, 102(2):342–346, 2008.

[125] RA Marcus. On the theory of oxidation-reduction reactions involving electron transfer. i. *J. Chem. Phys.*, 24:966, 1956.

[126] RA Marcus. Chemical and electrochemical electron-transfer theory. *Annu. Rev. Phys. Chem.*, 15(1):155–196, 1964.

[127] RA Marcus and N Sutin. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta*, 811(3):265–322, 1985.

[128] S. A De Marothy, M. R. A Blomberg, and P. E. M Siegbahn. Elucidating the mechanism for the reduction of nitrite by copper nitrite reductase—a contribution from quantum chemical studies. *Journal of Computational Chemistry*, 28(2):528–539, Jan 2006.

[129] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, April 2004.

[130] J Meiler and D Baker. Rosettaligand: Protein small-molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics*, 65(3):538–548, 2006.

[131] P Melchiorri and L Negri. The dermorphin peptide family. *General Pharmacology: The Vascular System*, 27(7):1099–1107, 1996.

[132] P.C MONTECUCCHI, ADA ANASTASI, R CASTIGLIONE, and V ERSPAMER. Isolation and amino acid composition of sauvagine. *International Journal of Peptide and Protein Research*, 16(3):191–199, 1980.

[133] P.C MONTECUCCHI, R CASTIGLIONE, S PIANI, L GOZZINI, and V ERSPAMER. Amino acid composition and sequence of dermorphin, a novel opiate-like peptide from the skin of phyllomedusa sauvagei. *International Journal of Peptide and Protein Research*, 17(3):275–283, 1981.

[134] P.M Murphy, J.M Bolduc, J.L Gallaher, B.L Stoddard, and D Baker. Alteration of enzyme specificity by computational loop remodeling and design. *Proceedings of the National Academy of Sciences*, 106(23):9215, 2009.

[135] KG Nicholson, FY Aoki, A Osterhaus, S Trottier, O Carewicz, CH Mercier, A Rode, N Kinnersley, and P Ward. Efficacy and safety of oseltamivir in treatment of acute influenza: a randomised controlled trial. *The Lancet*, 355(9218):1845–1850, 2000.

[136] K Nilsson, D Lecerof, E Sigfridsson, and U Ryde. An automatic method to generate force-field parameters for hetero-compounds. *Acta Crystallographica Section D: Biological Crystallography*, 59(2):274–289, 2003.

[137] J. K Norskov, F Abild-Pedersen, F Studt, and T Bligaard. Surface chemistry special feature: Density functional theory in surface chemistry and catalysis. *Proceedings of the National Academy of Sciences*, 108(3):937–943, Jan 2011.

[138] K Olesen, A Veselov, Y Zhao, Y Wang, B Danner, C.P Scholes, and J.P Shapleigh. Spectroscopic, kinetic, and electrochemical characterization of heterologously expressed wild-type and mutant forms of copper-containing nitrite reductase from rhodobacter sphaeroides 2.4. 3. *Biochemistry*, 37(17):6086–6094, 1998.

[139] M.H.M Olsson and U Ryde. The influence of axial ligands on the reduction potential of blue copper proteins. *Journal of Biological Inorganic Chemistry*, 4(5):654–663, 1999.

[140] G.A Omburo, J.M Kuo, L.S Mullins, and FM Raushel. Characterization of the zinc binding site of bacterial phosphotriesterase. *Journal of Biological Chemistry*, 267(19):13278, 1992.

[141] Alexey Onufriev, Donald Bashford, and David A Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics*, 55(2):383–394, Mar 2004.

[142] L Otvos. Peptide-based drug design: here and now. *Pept-Based Drug Des*, 494:1–8, 2008.

[143] W.D.M Paton. The action of morphine and related substances on contraction and on acetylcholine output of coaxially stimulated guinea-pig ileum. *British journal of pharmacology and chemotherapy*, 12(1):119, 1957.

[144] L. PAULING. Nature of forces between large molecules of biological interest. *Nature*, 161:707–709, May 1948.

[145] R.G Pearson. Hard and soft acids and bases, hsab, part 1: Fundamental principles. *Journal of Chemical Education*, 45(9):581, 1968.

[146] John P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, 33(12):8822–8824, Jun 1986.

[147] J.P Perdew, S Kurth, A Zupan, and P Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Physical Review Letters*, 82(12):2544–2547, 1999.

[148] J.P Perdew, J Tao, V.N Staroverov, and G.E Scuseria. Meta-generalized gradient approximation: Explanation of a realistic nonempirical density functional. *J. Chem. Phys.*, 120:6898, 2004.

[149] K Pierloot, J.O.A De Kerpel, U Ryde, M.H.M Olsson, and B.O Roos. Relation between the structure and spectroscopic properties of blue copper proteins. *Journal of the American Chemical Society*, 120(50):13156–13166, 1998.

[150] A.L Pinto, H.W Hellinga, and J.P Caradonna. Construction of a catalytically active iron superoxide dismutase by rational protein design. *Proceedings of the National Academy of Sciences of the United States of America*, 94(11):5562, 1997.

[151] P.F Predki, V Agrawal, A.T Brünger, and L Regan. Amino-acid substitutions in a surface turn modulate protein stability. *Nat Struct Mol Biol*, 3(1):54–58, 1996.

[152] M Prudêncio, RR Eady, and G Sawers. Catalytic and spectroscopic analysis of blue copper-containing nitrite reductase mutants altered in the environment of the type 2 copper centre: implications for substrate interaction. *Biochemical Journal*, 353(Pt 2):259, 2001.

[153] TD Rae, PJ Schmidt, RA Pufahl, VC Culotta, and T V O'Halloran. Undetectable intracellular free copper: the requirement of a copper chaperone for superoxide dismutase. *Science*, 284(5415):805, 1999.

[154] S Raman, O.F Lange, P Rossi, M Tyka, X Wang, J Aramini, G Liu, T.A Ramelot, A Eletsky, and T Szyperski. Nmr structure determination for larger proteins using backbone-only data. *Science*, 327(5968):1014, 2010.

[155] DW Randall, DR Gamelin, LB LaCroix, and EI Solomon. Electronic structure contributions to electron transfer in blue cu and cu a. *Journal of Biological Inorganic Chemistry*, 5(1):16–29, 2000.

[156] C.J Reedy and B.R Gibney. Heme protein assemblies. *Chemical Reviews*, 104(2):617–650, 2004.

[157] C.A Rohl, C.E.M Strauss, K Misura, and D Baker. Protein structure prediction using rosetta. *Methods in enzymology*, 383:66–93, 2004.

[158] C Roodveldt and D.S Tawfik. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry*, 44(38):12728–12736, 2005.

[159] D Röthlisberger, O Khersonsky, AM Wollacott, L Jiang, J DeChancie, J Betker, JL Gallaher, EA Althoff, A Zanghellini, and O Dym. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–195, 2008.

[160] L Rulisek and J Vondrasek. Coordination geometries of selected transition metal ions (co2+, ni2+, cu2+, zn2+, cd2+, and hg2+) in metalloproteins. *Journal of inorganic biochemistry*, 71(3-4):115–127, 1998.

[161] J.P Ryckaert, G Ciccotti, and H.J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[162] U Ryde. The coordination of the catalytic zinc ion in alcohol dehydrogenase studied by combined quantum-chemical and molecular mechanics calculations. *Journal of Computer-Aided Molecular Design*, 10(2):153–164, 1996.

[163] L.G Rydén and L.T Hunt. Evolution of protein complexity: The blue copper-containing oxidases and related proteins. *Journal of molecular evolution*, 36(1):41–66, 1993.

[164] M Sakurai, T Furuki, and Y Inoue. The pka of the zinc-bound water in carbonic anhydrase and its model compounds as studied by the am1 calculation coupled with a reaction field theory. *The Journal of Physical Chemistry*, 99(50):17789–17794, 1995.

[165] E. N Salgado, X. I Ambroggio, J. D Brodin, R. A Lewis, B Kuhlman, and F. A Tezcan. From the cover: Metal templated design of protein interfaces. *Proceedings of the National Academy of Sciences*, 107(5):1827–1832, Feb 2010.

[166] A Schafer, H Horn, and R Ahlrichs. Fully optimized contracted gaussian basis sets for atoms li to kr. *J. Chem. Phys*, 97:2571, 1992.

[167] Ansgar Schafer, Christian Huber, and Reinhart Ahlrichs. Fully optimized contracted gaussian basis sets of triple zeta valence quality for atoms li to kr. *The Journal of Chemical Physics*, pages 1–7, Jul 1999.

[168] Ansgar Schäfer, Andreas Klamt, Diana Sattel, John C. W Lohrenz, and Frank Eckert. Cosmo implementation in turbomole: Extension of an efficient quantum chemical code towards liquid systems. *Physical Chemistry Chemical Physics*, 2(10):2187–2193, Jan 2000.

[169] P.W Schiller, N.T.M Dung, N.N Chung, and C Lemieux. Dermorphin analogs carrying an increased positive net charge in their" message" domain display extremely high. mu.-opioid receptor selectivity. *J. Med. Chem.*, 32(3):698–703, 1989.

[170] JM Seminario. Calculation of intramolecular force fields from second-derivative tensors. *International Journal of Quantum Chemistry*, 60(7):1271–1277, 1996.

[171] D Shaw, P Maragakis, K Lindorff-Larsen, and S Piana. Atomic-level characterization of the structural dynamics of proteins. *Science*, Jan 2010.

[172] D.E Shaw, R.O Dror, J.K Salmon, JP Grossman, K.M Mackenzie, J.A Bank, C Young, M.M Deneroff, B Batson, and K.J Bowers. Millisecond-scale molecular dynamics simulations on anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, page 65, 2009.

[173] K Shibata, T Suzawa, S Soga, T Mizukami, K Yamada, N Hanai, and M Yamasaki. Improvement of biological activity and proteolytic stability of peptides by coupling with a cyclic peptide. *Bioorganic & medicinal chemistry letters*, 13(15):2583–2586, 2003.

[174] P Siddarth and RA Marcus. Comparison of experimental and theoretical electronic matrix elements for long-range electron transfer. *Journal of Physical Chemistry*, 94(7):2985–2989, 1990.

[175] V Sideraki, K.A Mohamedali, D.K Wilson, Z Chang, R.E Kellems, F.A Quiocho, and F.B Rudolph. Probing the functional role of two conserved active site aspartates in mouse adenosine deaminase. *Biochemistry*, 35(24):7862–7872, 1996.

[176] JB Siegel, A Zanghellini, HM Lovick, G Kiss, AR Lambert, JL St Clair, JL Gallaher, D Hilvert, MH Gelb, and BL Stoddard. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*, 329(5989):309, 2010.

[177] D Sindhikara, Y Meng, and A.E Roitberg. Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.*, 128:024103, 2008.

[178] J.F Smalley, S.W Feldberg, C.E.D Chidsey, M.R Linford, M.D Newton, and Y.P Liu. The kinetics of electron transfer through ferrocene-terminated alkanethiol monolayers on gold. *The Journal of Physical Chemistry*, 99(35):13141–13149, 1995.

[179] Edward I Solomon and Ryan G Hadt. Recent advances in understanding blue copper proteins. *Coordination Chemistry Reviews*, pages 1–16, Jan 2011.

[180] E.I Solomon, M.J Baldwin, and M.D Lowery. Electronic structures of active sites in copper proteins: Contributions to reactivity. *Chemical Reviews*, 92(4):521–542, 1992.

[181] E.I Solomon, R.K Szilagyi, S.D.B George, and L Basumallick. Electronic structures of metal sites in proteins and models: contributions to function in blue copper proteins. *Chemical Reviews*, 104(2):419–458, 2004.

[182] D.A Spivak, T.Z Hoffman, A.H Moore, M.J Taylor, and K.D Janda. A comparison of flexible and constrained haptens in eliciting antibody catalysts for paraoxon hydrolysis. *Bioorganic & medicinal chemistry*, 7(6):1145–1150, 1999.

[183] V.N Staroverov, G.E Scuseria, J Tao, and J.P Perdew. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.*, 119:12129, 2003.

[184] Claudia Steffen, Klaus Thomas, Uwe Huniar, Arnim Hellweg, Oliver Rubner, and Alexander Schroer. Tmolex-a graphical user interface for turbomole. *Journal of Computational Chemistry*, 2010.

[185] PJ Stephens, FJ Devlin, CF Chabalowski, and MJ Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, 1994.

[186] M Sundararajan, I.H Hillier, and N.A Burton. Mechanism of nitrite reduction at t2cu centers: Electronic structure calculations of catalysis by copper nitrite reductase and by synthetic model compounds. *The Journal of Physical Chemistry B*, 111(19):5511–5517, 2007.

[187] R.H Swendsen and J.S Wang. Replica monte carlo simulation of spin-glasses. *Physical Review Letters*, 57(21):2607–2609, 1986.

[188] T Terasaka, K Tsuji, T Kato, I Nakanishi, T Kinoshita, Y Kato, M Kuno, T Inoue, K Tanaka, and K Nakamura. Rational design of non-nucleoside, potent, and orally bioavailable adenosine deaminase inhibitors: predicting enzyme conformational change and metabolism. *J. Med. Chem.*, 48(15):4750–4753, 2005.

[189] F Toma, V Dive, S Fermandjian, K Darlak, and Z Grzonka. Preferred solution and calculated conformations of dermorphin and analysis of structure-conformation-activity relationships in the series [ala^n]-dermorphin. *Biopolymers*, 24(12):2417–2430, 1985.

[190] A Tramontano, A.A Ammann, and R.A Lerner. Antibody catalysis approaching the activity of enzymes. *Journal of the American Chemical Society*, 110(7):2282–2286, 1988.

[191] A Troisi and G Orlandi. The hole transfer in dna: calculation of electron coupling between close bases. *Chemical Physics Letters*, 344(5-6):509–518, 2001.

[192] B.L Vallee and D.S Auld. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry*, 29(24):5647–5659, 1990.

[193] B.L Vallee and F.L Hoch. Zinc, a component of yeast alcohol dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 41(6):327, 1955.

[194] J.L Vanhooke, M.M Benning, F.M Raushel, and H.M Holden. Three-dimensional structure of the zinc-containing phosphotriesterase with the bound substrate analog diethyl 4-methylbenzylphosphonate. *Biochemistry*, 35(19):6020–6025, 1996.

[195] L.K Vaughn, W.S Wire, P Davis, Y Shimohigashi, G Toth, R.J Knapp, V.J Hruby, T.F Burks, and H.I Yamamura. Differentiation between rat brain and mouse vas deferens [delta] opioid receptors. *European journal of pharmacology*, 177(1-2):99–101, 1990.

[196] J Villa and A Warshel. Energetics and dynamics of enzymatic reactions. *The Journal of Physical Chemistry B*, 105(33):7887–7907, 2001.

[197] C.L Vizcarra and S.L Mayo. Electrostatics in computational protein design. *Current Opinion in Chemical Biology*, 9(6):622–626, 2005.

[198] J Wang, W Wang, P.A Kollman, and D.A Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling*, 25(2):247–260, 2006.

[199] A Warshel. Electrostatic basis of structure-function correlation in proteins. *Accounts of chemical research*, 14(9):284–290, 1981.

[200] A Warshel, P.K Sharma, M Kato, Y Xiang, H Liu, and M.H.M Olsson. Electrostatic basis for enzyme catalysis. *Chemical Reviews*, 106(8):3210–3235, 2006.

[201] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297, Jan 2005.

[202] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297, Jan 2005.

[203] H.J Wijma, I MacPherson, O Farver, E.I Tocheva, I Pecht, M.P Verbeet, M.E.P Murphy, and G.W Canters. Effect of the methionine ligand on the reorganization energy of the type-1 copper site of nitrite reductase. *Journal of the American Chemical Society*, 129(3):519–525, 2007.

[204] D.K Wilson and F.A Quiocho. A pre-transition-state mimic of an enzyme: X-ray structure of adenosine deaminase with bound 1-deazaadenosine and zinc-activated water. *Biochemistry*, 32(7):1689–1694, 1993.

[205] KY Wong and J Gao. The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined qm/mm simulations. *Biochemistry*, 46(46):13352–13369, 2007.

[206] M.W Wong. Vibrational frequency prediction using density functional theory. *Chemical Physics Letters*, 256(4-5):391–399, 1996.

[207] C Yanover, M Fromer, and JM Shifman. Dead-end elimination for multistate protein design. *Journal of Computational Chemistry*, 28(13):2122–2129, 2007.

[208] H Zhang, M.J Boulanger, A.G Mauk, and M.E.P Murphy. Carbon monoxide binding to copper-containing nitrite reductase from alcaligenes faecalis. *The Journal of Physical Chemistry B*, 104(46):10738–10742, 2000.

[209] R Zhou and B.J Berne. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12777, 2002.